

Randomized Solutions to Convex Programs with Multiple Chance Constraints*

Georg Schildbach[†]

Lorenzo Fagiano[‡]

Manfred Morari[†]

March 3, 2013

Abstract

The scenario-based optimization approach ('scenario approach') provides an intuitive way of approximating the solution to chance-constrained optimization programs, based on finding the optimal solution under a finite number of sampled outcomes of the uncertainty ('scenarios'). A key merit of this approach is that it neither assumes knowledge of the uncertainty set, as it is common in robust optimization, nor of its probability distribution, as it is usually required in stochastic optimization. Moreover, the scenario approach is computationally efficient as its solution is based on a deterministic optimization program that is canonically convex, even when the original chance-constrained problem is not. Recently, researchers have obtained theoretical foundations for the scenario approach, providing a direct link between the number of scenarios and bounds on the constraint violation probability.

These bounds are tight in the general case of an uncertain optimization problem with a single chance constraint. However, this paper shows that these bounds can be improved in situations where the constraints have a limited 'support rank', a new concept that is introduced for the first time. This property is typically found in a large number of practical applications—most importantly, if the problem originally contains multiple chance constraints (e.g. multi-stage uncertain decision problems), or if a chance constraint belongs to a special class of constraints (e.g. linear or quadratic constraints). In these cases the quality of the scenario solution is improved while the same bound on the constraint violation probability is maintained, and also the computational complexity is reduced.

Key words: Uncertain Optimization, Chance Constraints, Randomized Methods, Convex Optimization, Scenario Approach, Multi-Stage Decision Problems.

1 Introduction

Optimization is ubiquitous in modern problems of engineering and science, where a decision or design variable $x \in \mathbb{R}^d$ has to be selected from a constrained set $\mathcal{X} \subseteq \mathbb{R}^d$ and its quality is measured against some objective or cost function $f_0 : \mathbb{R}^d \rightarrow \mathbb{R}$. A common difficulty in these problems is the lack of precise information about some of the underlying data, so that for a particular solution it remains uncertain whether the constraint are satisfied and which objective or cost value is achieved. This uncertainty shall be denoted by an (unknown) abstract variable $\delta \in \Delta$, where Δ is understood as the uncertainty set of a non-specified nature. It may affect the target function f_0 and/or the feasible set \mathcal{X} , where the latter situation represents a particular challenge since good solutions are usually found close to the boundary of this set, which requires a trade-off between the objective value and the strictness of constraint satisfaction.

*This manuscript is the preprint of a paper submitted to the SIAM Journal on Optimization and it is subject to SIAM copyright. SIAM maintains the sole rights of distribution or publication of the work in all forms and media. If accepted, the copy of record will be available at <http://www.siam.org>.

[†]Automatic Control Laboratory, Swiss Federal Institute of Technology, Zurich, Switzerland (schildbach|morari@control.ee.ethz.ch).

[‡]Dipartimento di Automatica e Informatica, Politecnico di Torino, Torino, Italy and Department of Mechanical Engineering, University of California at Santa Barbara, Santa Barbara (CA), United States (lorenzo.fagiano@polito.it).

A large variety of approaches addressing this issue have been proposed in the areas of robust and stochastic optimization [3, 4, 6, 16, 17, 19, 21, 23], with the preferred method-of-choice depending on the particular problem at hand.

For many practical applications, the formulation of chance-constraints has proven to be the appropriate concept to handle uncertainty. Here the uncertainty δ is assumed to have a stochastic nature, and the decision variable x is allowed to fall outside of the feasible set \mathcal{X} , but only with a probability no higher than a given threshold $\varepsilon \in (0, 1)$. Yet chance-constrained optimization programs generally remain difficult to solve. In special cases, they can be tackled by stochastic optimization techniques, when also a probability distribution function for δ is assumed [6, 16, 21, 23]. However, their deterministic reformulation is usually non-convex, and it can be difficult even to find a feasible point with a good objective function value.

Recent contributions have shown that randomized algorithms are a viable approach for finding, with high confidence, a sub-optimal solution to chance-constrained programs; see [9–13] and the references therein. They propose a scenario-based optimization approach (*‘scenario approach’*) for finding values of x that are, with a very high confidence, feasible with respect to the chance constraints and achieve good objective function values. Their theory considers the following general class of problems:

$$\min_{x \in \mathbb{X}} c^T x, \quad (1.1a)$$

$$\text{s.t.} \quad \Pr\{\delta \in \Delta \mid x \in \mathcal{X}(\delta)\} \geq (1 - \varepsilon), \quad (1.1b)$$

where $c \in \mathbb{R}^d$ defines the linear objective function, $\mathbb{X} \subset \mathbb{R}^d$ represents a deterministic convex constraint set, and $\mathcal{X}(\delta) \subset \mathbb{R}^d$ is a convex set that depends on the uncertainty $\delta \in \Delta$, in which the decision variable has to lie with a probability of at least $1 - \varepsilon$. While (1.1) appears to be restrictive on first sight, it actually encompasses a vast range of problems—namely any uncertain optimization program which is convex whenever the value of the uncertainty δ is fixed.

The fundamental idea of the scenario approach is simple and intuitive: it draws a finite number $K \in \mathbb{N}$ of independent and identically distributed (i.i.d.) samples of the uncertainty δ , and proposes to use the solution of the optimization program (*‘scenario program’*) obtained by replacing the chance constraint (1.1b) with the resulting K sampled constraints, which are deterministic and convex. Strong results have been developed by [10, 12], providing tight bounds for the choice of K when linking it directly to the probability with which the solution to the scenario program (*‘scenario solution’*) violates the chance constraint (1.1b). Furthermore as shown by [10, 13], these bounds extend to the case where after sampling K constraints a given number of $R \in \mathbb{N}$ sampled constraints are removed, a procedure that can be applied to improve the objective value of the scenario solution.

A key advantage of the scenario approach is that it neither requires knowledge of the distribution function of the uncertainty nor of the uncertainty set Δ , which may be of a very general nature; it only assumes the availability of a sufficient number of i.i.d. random samples of δ . Therefore it could be argued that the scenario approach is at the heart of any robust and stochastic optimization method, because either the uncertainty set Δ or the probability distribution of δ are constructed based on some (necessarily finite) experience of the uncertainty. Another computational advantage is that the solution is obtained through a deterministic convex optimization program, for which numerous efficient algorithms and solvers are readily available [8, 18, 20].

While the existing theory on the scenario approach, and also in stochastic programming, can only handle single chance constraints, as in (1.1), problems with multiple chance constraints are also of interest:

$$\min_{x \in \mathbb{X}} c^T x, \quad (1.2a)$$

$$\text{s.t.} \quad \Pr\{\delta \in \Delta \mid x \in \mathcal{X}_i(\delta)\} \geq (1 - \varepsilon_i) \quad \forall i \in \mathbb{N}_1^N, \quad (1.2b)$$

where i is the chance constraint index running in $\mathbb{N}_1^N := \{1, 2, \dots, N\}$. One motivation for this is that systems in science and engineering, which are steadily increasing in size and complexity, often comprise multiple sub-systems that are individually subject to uncertain constraints, yet all sub-systems have to be considered integratively for an optimal operation of the entire system. Another motivation are multi-stage

stochastic decision problems [6, Cha. 7], [16, Cha. 8] [21, Cha. 13] [23, Cha. 3] in which typically one chance constraint needs to be enforced per time step, while the optimal decision problem comprises multiple time steps. With view on the latter, the N chance constraints of (1.2) will also be referred to as ‘*stages*’ in this paper; yet the multiple chance constraints are treated in a general, i.e. not necessarily temporal, manner.

Note that, in principle, a sub-optimal solution to (1.2) can be obtained by means of finding a sub-optimal solution to (1.1), if the a single chance constraint is set up as follows:

$$\mathcal{X} := \mathcal{X}_1 \cap \mathcal{X}_2 \cap \dots \cap \mathcal{X}_N \quad \text{and} \quad \varepsilon := \min\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N\} . \quad (1.3)$$

However, this procedure may introduce a considerable amount of conservatism, as it requires the scenario solution x to lie simultaneously inside *all* constraint sets \mathcal{X}_i with the *highest* of all probabilities $1 - \varepsilon_i$. Apparently, this conservatism increases for a larger number of chance constraints N and a greater variation in the values of ε_i .

The fundamental contribution of this paper is to reduce the existing sample bounds if the scenario approach is applied to stochastic programs with multiple chance constraints—with the key effects of an improvement of the quality of the scenario solution and a decrease of the computational complexity, while the probabilistic guarantees remain the exactly the same as before. The developed theory also applies to many single-stage problems, where the existing sample bounds (which are tight in the general case) are reduced in cases when the chance constraint carries a special property, as it is held by many common constraint classes. Intuitively speaking, the property is that the chance constraint fails to constrain some ‘directions’ of the search space, like a linear constraint with an uncertain right-hand side leaves a $(d - 1)$ -subspace of \mathbb{R}^d unconstrained.

The paper is organized as follows. Section 2 exhibits a technical statement of the problem, and Section 3 provides further background on its properties. Section 4 introduces the new concept of the ‘support rank’ of a chance constraint, which directly leads to the derivation of the main results of this paper: sample bounds for the multi-stage stochastic program are derived in Section 5, and a sampling-and-discarding procedure is described in Section 6. Finally, Section 7 presents an example that demonstrates the potential benefits of this theory in certain applications, when compared to the classic scenario approach.

2 Problem Formulation

2.1 Stochastic Program with Multiple Chance Constraints

Consider a stochastic optimization problem with linear objective function and multiple chance constraints, which shall be referred to as the *Multi-Stage Stochastic Program* MSP:

$$\min_{x \in \mathbb{X}} c^T x , \quad (2.1a)$$

$$\text{s.t.} \quad \Pr[f_i(x, \delta) \leq 0] \geq (1 - \varepsilon_i) \quad \forall i \in \mathbb{N}_1^N . \quad (2.1b)$$

Here $c \in \mathbb{R}^d$ defines a linear target function, and $\mathbb{X} \subset \mathbb{R}^d$ represents a compact and convex set, which can be regarded as the combination of any set of deterministic convex constraints. It also fulfills the technical purpose of ensuring the boundedness of any solution to (2.1). \Pr denotes a probability measure on the uncertainty set Δ , with respect to the random variable $\delta \in \Delta$. The *chance constraints* $i \in \mathbb{N}_1^N$ in (2.1b) will be referred to as *stages* of the MSP, in order to avoid ambiguities in the nomenclature of this paper. They involve *constraint functions* $f_i : \mathbb{R}^d \times \Delta \rightarrow \mathbb{R}$ that depend on the optimization variable $x \in \mathbb{R}^d$ as well as on the random uncertainty $\delta \in \Delta$ in the following manner.

Assumption 2.1 (Convexity) *The constraint functions $f_i(\cdot, \cdot)$ of every stage $i \in \mathbb{N}_1^N$ satisfy the property that for almost every uncertainty $\delta \in \Delta$, $f_i(\cdot, \delta)$ is a convex function in the optimization variable.*

Except for Assumption 2.1, the constraint functions $f_i(\cdot, \cdot)$ are entirely arbitrary; in particular, their dependence on the random variables δ is completely generic.

The inequality ' $f_i(x, \delta) \leq 0$ ' in a chance constraint of (2.1b) is referred to as the *nominal chance constraint*, while $\varepsilon_i \in (0, 1)$ represents the *chance constraint level*. For any given $x \in \mathbb{R}^d$, the nominal chance constraint characterizes the subset of Δ whose complement needs to be controlled (in its probability measure) by ε_i , in order for x to be considered as a feasible point of the particular stage i . This becomes more obvious by noting that the probability notation in (2.1b) stands short for

$$\Pr[f_i(x, \delta) \leq 0] := \Pr\{\delta \in \Delta \mid f_i(x, \delta) \leq 0\} , \quad (2.2)$$

avoiding the cumbersome way of writing out the probability as the measure of a subset of Δ . Both notations are used in the sequel, where the short notation (indicated by square brackets) is preferred for the sake of simplicity and the long notation (indicated by curly brackets) is used when it brings additional clarity to the expression.

The use of 'min' instead of 'inf' in (2.1a) is justified by the fact that the feasible set of a single stage can be shown to be closed for very general cases [16, Thm. 2.1]. The feasible set is then compact, by the presence of \mathbb{X} , and any infimum is indeed attained; otherwise 'inf' can always be replaced by 'min' for finding an optimal point in the closure of the feasible set.

The results presented in this paper are distribution-free in the sense that they do not rely on the knowledge of the measure \Pr , or the corresponding probability distribution of δ . It is required however that such a probability measure on Δ exists, so that \Pr can be used throughout this paper for analysis purposes. In fact, not even the support set Δ , which can be very general in nature needs, needs to be known.

Instead, it is assumed that a sufficient number of independent random samples can be obtained from δ . Moreover, it remains a standing assumption that the σ -algebra of \Pr -measurable sets in Δ is large enough to contain all sets whose probability measure is used in this paper, like the one in (2.2). These requirements appear to be reasonable for a large range of practical applications, see also [12, p. 4]

Remark 2.2 (Generality of Problem Formulation) *The formulation (2.1) comprises, in particular, the following problem settings. (Some of them actually appear in the example of Section 7.) (a) A random, non-linear, convex objective function can be included by an epigraph reformulation, with the new objective being a scalar and hence linear [8, Sec. 3.1.7]. (b) The constraint function of a vector-valued nominal chance constraint, also known as a joint chance constraint, can be transformed into a scalar, non-linear, convex constraint function by taking the point-wise maximum of its component functions, preserving its convexity [8, Sec. 3.2.3]. (c) If different stages depend on different uncertain variables, all uncertainties can be combined into the vector uncertainty δ , as not all stages have to depend explicitly on all entries of δ . (d) Any non-random and convex constraints can be readily included as part of the convex and compact set \mathbb{X} .*

The following assumption is made in order to avoid unnecessary technical issues, which are of little relevance in most practical applications; compare [12, Ass. 1].

Assumption 2.3 (Existence and Uniqueness) *(a) Problem (2.1) admits at least one feasible point. By the compactness of \mathbb{X} , this implies that there exists at least one optimal point of (2.1). (b) If there are multiple optimal points of (2.1), a unique one is selected by the help of an appropriate tie-break rule (e.g. the lexicographic order on \mathbb{R}^d).*

2.2 The Randomization Approach

From the basic literature on stochastic optimization, e.g. [6, 16, 21, 23], it is known that the MSP is generally a hard problem to solve. Even in the case when the joint distribution of δ is known, (2.1) is equivalent to the solution of a non-convex deterministic optimization program, except for a few very special cases. Furthermore, the probability calculations involve the computation of convolution integrals, which generally have to be approximated by numerical means.

The randomized optimization approach proposed in this paper can be used to obtain an approximate solution to the MSP, which is a feasible point of every stage $i = 1, \dots, N$ with a selected confidence probability $(1 - \theta_i)$. It is closely related to the theory on the scenario approach, as described e.g. in [10–13].

Suppose that for each stage $i \in \mathbb{N}_1^N$ a certain number of $K_i \in \mathbb{N}$ random samples (or ‘uncertainty scenarios’ [11]) is drawn for the uncertainty δ . They are denoted by $\delta^{(i, \kappa_i)}$ for $\kappa_i \in \{1, \dots, K_i\}$ and, for brevity of notation, also as collective *multi-samples* $\omega^{(i)} := \{\delta^{(i, 1)}, \dots, \delta^{(i, K_i)}\}$. The collection of all samples shall be combined into the multi-sample $\omega := \{\omega^{(1)}, \dots, \omega^{(N)}\}$, with their total number given by $K := \sum_{i=1}^N K_i$. All of these samples should be considered as ‘identical copies’ of the random uncertainty δ , in the sense that they are themselves random variables and satisfy the following key assumption.

Assumption 2.4 (Independence and Identical Distribution) *The sampling procedure is designed such that the set of all random samples together with the actual uncertainty,*

$$\bigcup_{i \in \mathbb{N}_1^N} \{\delta^{(i, 1)}, \dots, \delta^{(i, K_i)}\} \cup \{\delta\} ,$$

constitutes a set of independent and identically distributed (i.i.d.) random variables.

The *Multi-Stage Randomized Program* $\text{MRP}[\omega^{(1)}, \dots, \omega^{(N)}]$ is constructed as follows:

$$\min_{x \in \mathbb{X}} c^T x , \tag{2.3a}$$

$$\text{s.t. } f_i(x, \delta^{(i, \kappa_i)}) \leq 0 \quad \forall \kappa_i \in \mathbb{N}_1^{K_i}, i \in \mathbb{N}_1^N . \tag{2.3b}$$

Observe that in problem (2.3) the target function of (2.1a) is minimized while enforcing each nominal chance constraint for all values in its corresponding random multi-sample. Thus problem (2.3) is stochastic in the sense that its solution depends on the random samples in ω . It is important only for analytic purposes, while it is actually solved for the observations of the random samples, leading to the deterministic version of the Multi-Stage Randomized Program $\overline{\text{MRP}}[\bar{\omega}^{(1)}, \dots, \bar{\omega}^{(N)}]$:

$$\min_{x \in \mathbb{X}} c^T x , \tag{2.4a}$$

$$\text{s.t. } f_i(x, \bar{\delta}^{(i, \kappa_i)}) \leq 0 \quad \forall \kappa_i \in \mathbb{N}_1^{K_i}, i \in \mathbb{N}_1^N . \tag{2.4b}$$

Problem (2.4) arises from (2.3) through replacement of the (random) *samples* $\delta^{(i, \kappa_i)}$, $\omega^{(i)}$, and ω by their (deterministic) *outcomes* $\bar{\delta}^{(i, \kappa_i)}$, $\bar{\omega}^{(i)}$, and $\bar{\omega}$. Throughout the paper, the outcomes are distinguished from their random variables by a bar. Due to Assumption (2.1), problem (2.4) is a convex program of a pre-known structure. As such it can be solved efficiently by choice of a suitable algorithm from convex optimization, see e.g. [8, 18, 20].

Remark 2.5 (Relation to the Scenario Approach) (a) Problem (2.3) is a generalization of a ‘random convex program’ [10] or a ‘scenario program’ [12]: instead of only a single stage ($N = 1$), it may contain multiple stages ($N \geq 1$). This can be relevant to applications involving several independent chance constraints, resulting e.g. from different conditions on local sub-systems or time steps in a multi-stage decision problem. (b) Problem (2.4) provides the cost-optimal solution that is feasible with respect to all nominal constraints $i = 1, \dots, N$ in (2.1) under each of the scenarios in $\bar{\omega}^{(1)}, \dots, \bar{\omega}^{(N)}$. In this sense the proposed randomization method can be related to the classic scenario approach, however the scenario numbers K_i used in the problem may generally differ between the stages. (c) An approximate solution to (2.1) can also be found by a reformulation using only a single stage, and hence applying the existing theory for randomized programs (RPs). Here the nominal chance constraint is defined as the point-wise maximum of the functions $f_1(\cdot, \delta), \dots, f_N(\cdot, \delta)$ and the chance constraint level as the minimum of $\varepsilon_1, \dots, \varepsilon_N$. However, in general this approach comes at the price of an increased conservatism and therefore a higher cost of the solution [11, Sec. 2.1.2], as demonstrated also in the example of Section 7.

The fundamental goal of the randomization theory is to show that the solution to (2.3) has a limited *violation probability*, that is the probability of violating the nominal chance constraints (2.1b). This basic notion is recalled in the next section.

2.3 Randomized Solution and Violation Probability

In order to avoid unnecessary technical complications, the following assumption is introduced to ensure that there always exists a feasible solution to problem (2.4), similar to [12, p. 3].

Assumption 2.6 (Feasibility) (a) For any number of samples K_1, \dots, K_N , the randomized problem (2.3) admits a feasible solution almost surely. (b) For the sake of notational simplicity, any \Pr -null set for which (a) may not hold is assumed to be removed from Δ .

Feasibility can be taken for granted in the majority of practical problems; if it does not hold for a particular application, a generalization of the presented theory to include the infeasible case follows the as shown in [10].

Given the existence of a solution to (2.4), uniqueness follows by Assumption 2.1 and by carrying over the tie-break rule of Assumption 2.3(b), see [22, Thm. 10.1, 7.1]. Hence the *solution map*

$$\bar{x}^* : \Delta^K \rightarrow \mathbb{X} , \quad (2.5)$$

which returns as $\bar{x}^*(\bar{\omega}^{(1)}, \dots, \bar{\omega}^{(N)})$ the unique optimal point of $\overline{\text{MRP}}[\bar{\omega}^{(1)}, \dots, \bar{\omega}^{(N)}]$ for a given observation of the multi-samples $\{\bar{\omega}^{(1)}, \dots, \bar{\omega}^{(N)}\} \in \Delta^K$, is well-defined. The solution map can also be applied to problem $\text{MRP}[\omega^{(1)}, \dots, \omega^{(N)}]$, for which it is denoted by $x^* : \Delta^K \rightarrow \mathbb{X}$. Then $x^*(\omega^{(1)}, \dots, \omega^{(N)}) \in \mathbb{R}^d$ represents a random vector whose probability distribution is unknown; in fact it is a complicated function of the geometry of the problem and its parameters. This random vector will also be referred to as the *randomized solution*.

Based on a given choice for the sample sizes K_1, \dots, K_N , the goal is to obtain a bound on the probability with which the randomized solution $x^*(\omega^{(1)}, \dots, \omega^{(N)})$ violates the nominal chance constraints of the MSP for the actual uncertainty δ . Apparently, this involves two levels of randomness present in the problem: the first introduced by the random samples in ω , which determine the randomized solution; and the second introduced by the random uncertainty δ , which determines whether or not the randomized solution violates the nominal chance constraints in (2.1b). For this reason, the randomization approach of this paper is also called a *double-level of probability approach* [9, Rem. 2.3].

To highlight the two probability levels more clearly, suppose first that the multi-sample $\bar{\omega}$ has already been observed, so that the randomized solution $\bar{x}^*(\bar{\omega}^{(1)}, \dots, \bar{\omega}^{(N)})$ is fixed. Then for every stage $i = 1, \dots, N$ in (2.1b), the *ex-post violation probability* $\bar{V}_i(\bar{\omega}^{(1)}, \dots, \bar{\omega}^{(N)})$ is given by

$$\bar{V}_i(\bar{\omega}^{(1)}, \dots, \bar{\omega}^{(N)}) := \Pr\{\delta \in \Delta \mid f_i(\bar{x}^*(\bar{\omega}^{(1)}, \dots, \bar{\omega}^{(N)}), \delta) > 0\} . \quad (2.6)$$

In particular, each \bar{V}_i is a deterministic value in $[0, 1]$. Now, if the multi-sample ω has not yet been observed, the randomized solution $x^*(\omega^{(1)}, \dots, \omega^{(N)})$ is a random vector and so the *ex-ante violation probability*

$$V_i(\omega^{(1)}, \dots, \omega^{(N)}) := \Pr\{\delta \in \Delta \mid f_i(x^*(\omega^{(1)}, \dots, \omega^{(N)}), \delta) > 0\} . \quad (2.7)$$

becomes itself a random variable with support in $[0, 1]$, for every stage $i = 1, \dots, N$ in (2.1b). Hence the goal is to ensure that $V_i(\omega^{(1)}, \dots, \omega^{(N)}) \leq \varepsilon_i$ for all $i = 1, \dots, N$, with a sufficiently high confidence probability $(1 - \theta_i)$. Before these results can be derived however, some structural properties of the randomized program and technical lemmas ought to be discussed.

3 Structural Properties of the MRP

This section discusses some fundamental facts on RPs, partly based on existing theory [10–12] but also introduces some new concepts and results. This lays the technical groundwork that is required for the proof of the main results of this paper. Many of the concepts presented here are in their essence even more general, see [15], but this lies beyond the scope of this paper.

3.1 Support and Essential Constraint Sets

Similar to the solution map and violation probability, the new concepts of this section are defined first for the deterministic problem $\overline{\text{MRP}}$, and then carried over in a probabilistic sense to the stochastic problem MRP.

Definition 3.1 (Support Constraint Set) For some $i \in \mathbb{N}_1^N$ and $\kappa_i \in \mathbb{N}_1^{K_i}$, constraint $f_i(x, \bar{\delta}^{(i, \kappa_i)}) \leq 0$ is a support constraint of (2.4) if its removal from the problem entails a change in the optimal solution:

$$\bar{x}^*(\bar{\omega}^{(1)}, \dots, \bar{\omega}^{(N)}) \neq \bar{x}^*(\bar{\omega}^{(1)}, \dots, \bar{\omega}^{(i-1)}, \bar{\omega}^{(i)} \setminus \{\bar{\delta}^{(i, \kappa_i)}\}, \bar{\omega}^{(i+1)}, \dots, \bar{\omega}^{(N)}) .$$

In this case the sample $\bar{\delta}^{(i, \kappa_i)}$ is also said to ‘generate this support constraint’. The sample indices of all support constraints of (2.4) are combined into the support (constraint) sets $\text{Sc}_i \subset \mathbb{N}_1^{K_i}$ and $\text{Sc} := \bigcup_{i=1}^N \text{Sc}_i$.

Definition 3.1 could be stated equivalently in terms of the objective value [10, p. 3430]: a constraint is a support constraint if and only if the optimal target function value (or its preference by the tie-break rule) is strictly larger than when the constraint is removed.

In the stochastic setting of the MRP, a particular random sample $\delta^{(i, \kappa_i)}$ generating a support constraint becomes of course a random event, which can be associated with a certain probability. Similarly, the support constraint sets $\text{Sc}_1, \dots, \text{Sc}_N$ and Sc are naturally random sets.

While the removal of a single non-support constraint from $\overline{\text{MRP}}$ does not affect its solution, this does not mean that the non-support constraints can be omitted from the problem all at once without a change of the optimal point. For example, the configuration in Figure 3.1(c) includes only one support constraint, but the solution would change if the other constraints were removed all at once. This basic fact is the motivation for the definition of an *essential set*.

Definition 3.2 (Essential Constraint Set) An essential (constraint) set of $\overline{\text{MRP}}[\bar{\omega}_1, \dots, \bar{\omega}_N]$ is a subset of the constraints in (2.4b), generated by some $\{\tilde{\omega}_1, \dots, \tilde{\omega}_N\} \subseteq \{\bar{\omega}_1, \dots, \bar{\omega}_N\}$, for which the following two conditions hold: (a) the solution of the reduced problem remains the same as that of the full problem, i.e.

$$x^*(\tilde{\omega}_1, \dots, \tilde{\omega}_N) = x^*(\bar{\omega}^{(1)}, \dots, \bar{\omega}^{(N)}) ;$$

(b) all of the constraints in the essential set are support constraints of the reduced problem $\overline{\text{MRP}}[\tilde{\omega}_1, \dots, \tilde{\omega}_N]$.

To be precise, Definition 3.2 should also allow for the set \mathbb{X} to be an element of the essential set. However, this can be considered as a minor subtlety which is assumed to be understood in the sequel.

Unlike the support constraint set Sc , the essential constraint set of the deterministic problem $\overline{\text{MRP}}$ is not necessarily unique; in this case the problem is said to be *degenerate*. If the problem is *non-degenerate*, then its (unique) essential set is exactly equal to Sc , see [10, Def. 2.7]. For example, the configuration depicted in Figure 3.1(c) is degenerate as it contains two different essential sets. The stochastic problem MRP is called *non-degenerate* if it is non-degenerate with probability one, and *degenerate* otherwise.

Suppose that the MRP is non-degenerate and consider the cardinality $\text{card}(\text{Sc})$ of the support set Sc , which is a random integer in \mathbb{N}_0^d by [11, Thm. 2]. Then problem (2.3) is said to be *regular* if $\text{card}(\text{Sc})$ equals to some constant value $z \in \mathbb{N}_0^d$ with probability one. So if problem (2.3) is non-degenerate and *non-regular*, this means that it has a unique essential set equal to the support set with probability one, even though its cardinality takes on at least two different values with non-zero probability. An example of this situation is found in Figure 3.1, supposing that both situations (a) and (b) can occur with non-zero probability.

If degeneracy and non-regularity are simultaneously present in MRP, then the above concept of regularity is adjusted by considering the *minimal* cardinality of all essential sets. In other words, problem (2.3) is degenerate and non-regular if the minimal cardinality of all essential sets takes on at least two different values with non-zero probability. It is straightforward to imagine such a case with help of the following Example 3.3.

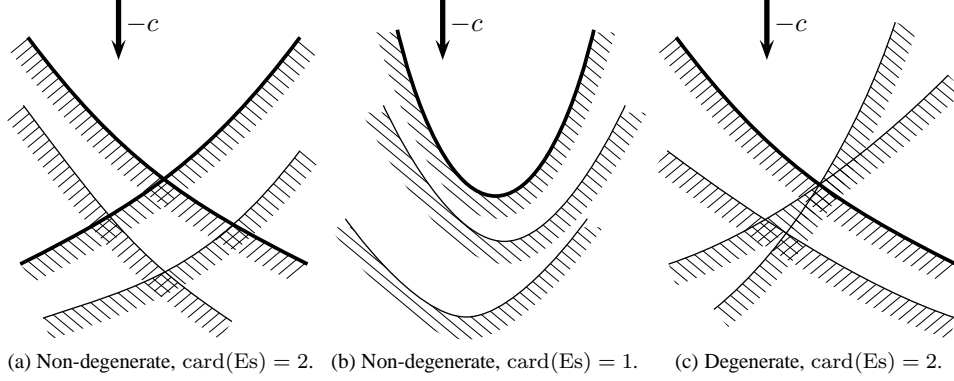


Figure 3.1: Illustration of Example 3.3 in \mathbb{R}^2 . The arrow represents the direction of lower target function values $-c$, bold lines indicate the support constraints of the respective configuration.

Example 3.3 For an illustration the above concepts, consider a multi-stage RP (2.3) in dimension $d = 2$, for which Figure 3.1 shows three configurations of sampled constraints.

The problem is non-degenerate and regular if its constraints are in a generic configuration, as sketched in either Figure 3.1(a) and (b), with probability one. This means that with probability one there are exactly two support constraints (or there is exactly one support constraint), which also form(s) the unique essential set.

Degeneracy means that a situation as in Figure 3.1(c) can occur with non-zero probability, where the essential set fails to be uniquely determined. Here there are two essential sets, both containing the support constraint and both having the cardinality of two.

Non-regularity, under non-degeneracy, means that the essential set is unique with probability one, but it can be of varying cardinality. For instance, if the two configurations of Figure 3.1(a,b) can both occur with positive probability, then the problem is non-regular and non-degenerate.

Finally, the problem is degenerate and non-regular if both phenomena can occur with a non-zero probability. This concludes Example 3.3, and naturally leads to a closer examination of the consequences of degeneracy and non-regularity.

3.2 Degeneracy and Non-Regularity

Both degeneracy and non-regularity play an important role in the derivation of the main results of this paper, and therefore need to be examined more closely. Note that it seems unrealistic to rule out either possibility by assumption, because their presence in very simple examples suggests that they do occur in a variety of practical applications.

Previous results for the single-stage RP case (where $N = 1$) start with the simplifying assumption that problem (2.3) is non-degenerate and regular with probability one. In another step, this assumption is then relaxed in order to include the exceptional cases, which adds a significant level of complexity to the arguments. Leveraging these insightful arguments in [10, 12] (for $N = 1$), for the sake of brevity degeneracy and non-regularity shall be dealt with in an integrative way. Some preliminary technical results are required for this and stated below.

Degeneracy

The authors of [12] propose a ‘heating-cooling’ procedure transforming the degenerate constraint system into a non-degenerate system (and back), where both systems are shown to have the same constraint violation characteristics. A similar argument is used in [10], relying on existing infinitesimal perturbation

techniques in order to recover a (arbitrarily close) non-degenerate constraint system from the degenerate one.

Similarly to the cited procedures, the approach chosen here derives a non-degenerate problem from the degenerate one, but in a slightly different manner. Namely, a unique essential set is selected by the introduction of another tie-break rule, used for determining whether or not a constraint is considered to be essential. Concretely, this means that in addition to its generating random sample $\delta^{(i, \kappa_i)}$, every constraint in (2.3b) has an associated tie-break random variable which is, for instance, uniformly distributed in $[0, 1]$. When two (or more) different essential constraint sets are compared, a unique one can be selected with probability one, based on the lowest sum of tie-break values.

Definition 3.4 (Minimal Essential Set) *The minimal essential (constraint) set $\overline{\text{Es}}$ of $\overline{\text{MRP}}$ is the (with probability one) unique essential set that (a) has the minimal possible cardinality and (b) has the lowest sum of tie-break values among all minimal essential sets.*

To be precise, in Definition 3.4 a tie-break random variable is assigned to the set \mathbb{X} as well, which may also be included in the minimal essential set besides the sampled constraints of (2.4b). Hence the uniqueness of the minimal essential set $\overline{\text{Es}}$, by virtue of the tie-breaker, allows to refer to its elements as the *essential constraints* of problem (2.4).

The following lemma is an adaptation of the Sampling Lemma in [15, Lem. 1.1] for the purposes of multi-stage RPs.

Lemma 3.5 (Sampling Lemma) *Suppose another sampled constraint of stage $i \in \mathbb{N}_1^N$, for the sample $\bar{\delta}$, is added to the problem $\overline{\text{MRP}}[\bar{\omega}^{(1)}, \dots, \bar{\omega}^{(N)}]$:*

$$f_i(x, \bar{\delta}) \leq 0 . \quad (3.1)$$

If the solution $\bar{x}^(\bar{\omega}^{(1)}, \dots, \bar{\omega}^{(N)})$ violates this additional sampled constraint, then the latter must be an essential constraint of the extended problem $\overline{\text{MRP}}[\bar{\omega}^{(1)}, \dots, \bar{\omega}^{(i-1)}, \bar{\omega}^{(i)} \cup \{\bar{\delta}\}, \bar{\omega}^{(i+1)}, \dots, \bar{\omega}^{(N)}]$.*

Proof. It is an immediate consequence of Definition 3.1 that the solution $\bar{x}^*(\bar{\omega}^{(1)}, \dots, \bar{\omega}^{(N)})$ violates the additional sampled constraint (3.1) if and only if the latter is a support constraint of the augmented problem $\overline{\text{MRP}}[\bar{\omega}^{(1)}, \dots, \bar{\omega}^{(i-1)}, \bar{\omega}^{(i)} \cup \{\bar{\delta}\}, \bar{\omega}^{(i+1)}, \dots, \bar{\omega}^{(N)}]$.

But if the new constraint (3.1) is a support constraint of the augmented problem, then it is part of all essential constraint sets by [10, Lem. 2.10], and therefore it must be contained in $\overline{\text{Es}}$. \square

Non-Regularity

The approach of [12] tackles non-regularity by the introduction of a ‘ball solution’, with the goal of recovering the property of full support. In [10], non-regularity is covered only in a deterministic sense, namely as far as it can be incorporated in the definition of ‘Helly’s dimension’. In this paper, the issue of non-regularity is resolved by the introduction of a conditional probability measure, where the conditional event is a fixed cardinality of the essential set.

For this purpose however, first the notion of regularity (as defined in Section 3.1) has to be refined to account for the presence of multiple stages. Observe that the minimal essential set $\overline{\text{Es}}$ of (2.4) can be broken down into

$$\overline{\text{Es}} = \overline{\text{Es}}_{\mathbb{X}} \cup \overline{\text{Es}}_1 \cup \overline{\text{Es}}_2 \cup \dots \cup \overline{\text{Es}}_N , \quad \text{with } \overline{\text{Es}}_{\mathbb{X}} \in \{\emptyset, \mathbb{X}\} ,$$

where $\overline{\text{Es}}_1, \dots, \overline{\text{Es}}_N$ denote the essential constraints contributed by the corresponding stages $i = 1, \dots, N$.

In the stochastic setting of the MRP, the sets Es and $\text{Es}_1, \dots, \text{Es}_N$ are random, and hence their cardinalities $z := \text{card}(\text{Es})$ and $z_i := \text{card}(\text{Es}_i)$ are random integers. In the sense of the previous definition of regularity, the multi-stage RP is *regular* if z equals to some constant value almost surely; it shall be called *multi-regular* if for all $i \in \mathbb{N}_1^N$ the individual cardinalities z_i equal to some constant values almost surely.

The following lemma will be important for tackling problems that fail to satisfy the property of multi-regularity. Suppose in problem (2.3) the number of samples is varied for each stage, while the sample sizes of the other stages remain constant. For example, in the case of the first stage $i = 1$ the varying sample size is denoted by \tilde{K}_1 (for the purpose of distinction from the actual sample size K_1), while the sample sizes for the other stages retain their constant values K_2, \dots, K_N . The resulting total number of samples is then denoted by $\tilde{K} := \tilde{K}_1 + K_2 + \dots + K_N$, and the corresponding multi-samples by $\tilde{\omega}^{(1)}, \tilde{\omega}$ instead of $\omega^{(1)}, \omega$, respectively.

Lemma 3.6 (Multi-Regularity) *Consider problem $\text{MRP}[\omega^{(1)}, \dots, \omega^{(i-1)}, \tilde{\omega}^{(i)}, \omega^{(i+1)}, \dots, \omega^{(N)}]$ for varying sizes \tilde{K}_i of the multi-sample $\tilde{\omega}^{(i)}$. For any $z_i \in \mathbb{N}$,*

$$\Pr^{\tilde{K}} \{ \tilde{\omega} \in \Delta^{\tilde{K}} \mid \text{card}(\text{Es}_i) = z_i \} \quad (3.2)$$

is either positive for all values of $\tilde{K}_i \geq z_i + 1$, or zero for all values of $\tilde{K}_i \geq z_i + 1$.

Proof. The result is shown for the first stage $i = 1$ only, whence it follows for all stages for reasons of symmetry.

Assume that the probability in (3.2) is positive for some $\tilde{K}_1 \geq z_1 + 1$; then with some non-zero probability there exists at least one sampled constraint which is not in the essential set Es_1 . Moreover, by virtue of Assumption 2.4, all of the \tilde{K}_1 sampled constraints are i.i.d. and thus none of them can be any more or less likely than another to become an essential constraint. So the probability that any specific $\tilde{K}_1 - 1$ out of the \tilde{K}_1 sampled constraints contain all z_1 essential constraints must be positive. Thus the probability of having an essential set of cardinality z_1 for $\tilde{K}_1 - 1$ samples is positive.

Following the above argument, the probability that any specific single sampled constraint out of the \tilde{K}_1 sampled constraint is not essential is non-zero. Since an additional $\tilde{K}_1 + 1$ -th sample is again i.i.d. to the previous \tilde{K}_1 samples, the probability that the $\tilde{K}_1 + 1$ -th sampled constraint is not essential is positive. Thus the probability of having an essential set of cardinality z_1 for $\tilde{K}_1 + 1$ samples is also positive.

The claimed result now follows by induction on \tilde{K}_1 : if the probability in (3.2) is positive for some $\tilde{K}_1 \geq z_1 + 1$, then it must be positive for all $\tilde{K}_1 \geq z_1$; consequently, if it is zero for some $\tilde{K}_1 \geq z_1 + 1$, then it must be zero for all $\tilde{K}_1 \geq z_1$ \square

4 Dimensions of the Multi-Stage Problem

The link between the sample sizes K_1, \dots, K_N and the corresponding violation probability of the randomized solution of (2.3) depends decisively on the ‘problem dimensions’. In the case of multi-stage RPs, the notion of the ‘problem dimension’ is more intricate than for the single-stage case—where it is simply determined by the maximum possible value for $\text{card}(\text{Es})$ and usually equals to the dimension d of the search space [11]. Not surprisingly, for multi-stage RPs the ‘dimension’ ceases to be a property of the problem as a whole; instead each of the stages has its own associated ‘dimension’.

4.1 The Support Dimension

The ‘dimension’ of a specific stage $i \in \mathbb{N}_1^N$ in the problem, embodied by the new concept of a *support dimension* (or *s-dimension*), is defined below.

Definition 4.1 (Support Dimension) *(a) The support dimension (or s-dimension) of a stage $i \in \mathbb{N}_1^N$ in (2.3) is the smallest integer ζ_i that satisfies*

$$\text{ess sup}_{\omega \in \Delta^K} \text{card}(\text{Es}_i) \leq \zeta_i \quad .$$

(b) Helly's dimension is the smallest integer ζ that satisfies

$$\text{ess sup}_{\omega \in \Delta^K} \text{card}(\text{Es}) \leq \zeta .$$

It should be pointed out that (finite) integers ζ and ζ_1, \dots, ζ_N matching Definition 4.1 always exist, so that the concepts of an s-dimension and Helly's dimension are indeed well-defined. This follows from the fact that there exists an upper bound on the total number of d support constraints in a (feasible and convex) d -dimensional optimization problem, e.g. [11, Thm. 2]. The latter result also provides immediate upper bounds on the s-dimensions, namely $\zeta_i \leq \zeta \leq d$ for every $i \in \mathbb{N}_1^N$.

In general, ζ_i represents an upper bound for $\text{card}(\text{Es}_i)$, which is used for determining the sample size K_i of stage i in Sections 5, 6. Yet in many practical problems the value of ζ_i may not be known exactly; in this case it has to be replaced by an upper bound ζ'_i of the actual s-dimension ζ_i . As the upper bound ζ'_i has, in fact, a positive influence on the sample size K_i , the tightness of this bound has a decisive impact on the solution quality (as suggested by Proposition 4.2 below). Helly's dimension provides a universal upper bound that is often easy to compute, but in many cases it can be significantly improved by exploiting structural properties of the constraints.

Proposition 4.2 (Probability Bound) *Consider some stage $i \in \mathbb{N}_1^N$ of $\text{MRP}[\omega^{(1)}, \dots, \omega^{(N)}]$, and let ζ'_i be an upper bound of its s-dimension ζ_i in the problem. Then the probability for any sample $\delta^{(i, \kappa_i)}$, for $\kappa_i \in \mathbb{N}_1^{K_i}$, of generating an essential constraint of (2.3) is bounded almost surely by*

$$\Pr^K \{ \omega \in \Delta^K \mid \kappa_i \in \text{Es} \} \leq \frac{\zeta'_i}{K_i} . \quad (4.1)$$

Proof. By virtue of Assumption 2.4 all samples in ω are independent, whence the event in (4.1) can be measured by the K -th product measure of \Pr , and also identically distributed, whence all constraints $\kappa_i \in \mathbb{N}_1^{K_i}$ generated for a fixed i in (2.3b) are probabilistically identical. Thus none of them can be any more or less likely than another to become an essential constraint.

The number of essential constraints $z_i := \text{card}(\text{Es}_i)$ is a random variable of unknown distribution. According to Definition 4.1(a), $z_i \leq \zeta_i$ almost surely, and by assumption $\zeta_i \leq \zeta'_i$. So consider any event in which $z_i = \bar{z}_i$, where \bar{z}_i is a constant value for which the event in (4.1) has non-zero probability. This last condition ensures that the conditional probability of a specific $\kappa_i \in \mathbb{N}_1^{K_i}$ generating an essential constraint, given that $z_i = \bar{z}_i$,

$$\Pr_{z_i = \bar{z}_i}^K \{ \omega \in \Delta^K \mid \kappa_i \in \text{Es} \} , \quad (4.2)$$

is well defined according to [5, Cha. 1.4] or [24, Cha. I.3]. For the notation of the conditional probability measure in (4.2), an index is added to the measure \Pr , indicating the conditional event ' $z_i = \bar{z}_i$ '.

Since by the first argument all K_i sampled constraints of stage i have the same probability of becoming an essential constraint, given that there are exactly \bar{z}_i essential constraints this probability equals to

$$\Pr_{z_i = \bar{z}_i}^K \{ \omega \in \Delta^K \mid \kappa_i \in \text{Es} \} = \frac{\bar{z}_i}{K_i} . \quad (4.3)$$

The claim follows from (4.3) and the fact that for any possible event the bound $z_i = \bar{z}_i \leq \zeta'_i$ holds. \square

4.2 The Support Rank

One possible property of a stage to rely on for upper bounds on its s-dimension proposed in this paper is the *support rank* (or *s-rank*) of the stage. In many problems it can be much tighter than Helly's dimension, in particular if the search space dimension d is large. Intuitively speaking, the s-rank of any stage $i \in \mathbb{N}_1^N$ in (2.3) is the dimension d of the search space \mathbb{R}^d less the maximal dimension of an *unconstrained subspace*, which is the maximal subspace that cannot be constrained by the sampled constraints of stage i , no matter the outcomes of the samples in $\omega^{(i)}$. In this respect, a very mild technical assumption is introduced that corresponds to similar conditions imposed for single-stage RPs [10, Thm. 3.3] and [12, Def. 3.1].

Assumption 4.3 The sample sizes K_1, \dots, K_N observe the lower bounds

$$K_i \geq \zeta_i + 1 \quad \forall i \in \mathbb{N}_1^N. \quad (4.4)$$

Before introducing the s-rank in a rigorous manner, three examples of constraint classes whose s-rank is easily bounded are given below. This equips the reader with the necessary intuition behind this concept, which can be used to determine the s-rank of a stage immediately for many practical problems.

Example 4.4 For each of the following cases, a visual illustration is provided in Figure 4.1.

(a) *Single Linear Constraint.* Suppose some stage $i \in \mathbb{N}_1^N$ in (2.1b) takes the linear form

$$f_i(x, \delta) = a^\top x - b(\delta), \quad (4.5)$$

where $a \in \mathbb{R}^d$ and $b : \Delta \rightarrow \mathbb{R}$ is any scalar that depends on the uncertainty. Then these constraints added to problem (2.3) for the samples $\delta^{(i,1)}, \dots, \delta^{(i,K_i)}$ are unable to constrain any direction in $\{a\}^\perp$, no matter the outcome of the multi-sample ω . Hence the s-rank of the stage in (4.5) equals to $\alpha = 1$.

(b) *Multiple Linear Constraints.* For a generalization of case (a), suppose that some stage $i \in \mathbb{N}_1^N$ in (2.1b) is given by

$$f_i(x, \delta) = A(\delta)^\top x - b(\delta), \quad (4.6)$$

where $A : \Delta \rightarrow \mathbb{R}^{r \times d}$ and $b : \Delta \rightarrow \mathbb{R}^r$ represent a matrix and a scalar that depend on the uncertainty δ . Moreover, suppose that the uncertainty enters the matrix $A(\delta)$ in such a way that the dimension of the linear span of its rows $A_{j,\cdot}(\delta)$, where $j = 1, \dots, r$, is known to satisfy

$$\dim \text{span}\{A_{j,\cdot}(\delta) \mid j \in \mathbb{N}_1^r, \delta \in \Delta\} \leq \beta < d.$$

Then these constraints added to problem (2.3) for the samples $\delta^{(i,1)}, \dots, \delta^{(i,K_i)}$ are unable to constrain any direction in the orthogonal subspace of the span, so the s-rank of the stage in (4.6) equals to β .

(c) *Quadratic Constraint.* For a nonlinear example, consider the case where some stage $i \in \mathbb{N}_1^N$ in (2.1b) is given by

$$f_i(x, \delta) = (x - x_c(\delta))^\top Q (x - x_c(\delta)) - r(\delta), \quad (4.7)$$

where $Q \in \mathbb{R}^{d \times d}$ is positive semi-definite with $\text{rank } Q = \gamma < d$, and $x_c : \Delta \rightarrow \mathbb{R}^d, r : \Delta \rightarrow \mathbb{R}_+$ represent a vector and scalar that depend on the uncertainty. Then these constraints added to problem (2.3) for the samples $\delta^{(i,1)}, \dots, \delta^{(i,K_i)}$ are unable to constrain any direction in the nullspace of the matrix Q , which has dimension $d - \gamma$, and so the s-rank of the stage in (4.7) equals to γ .

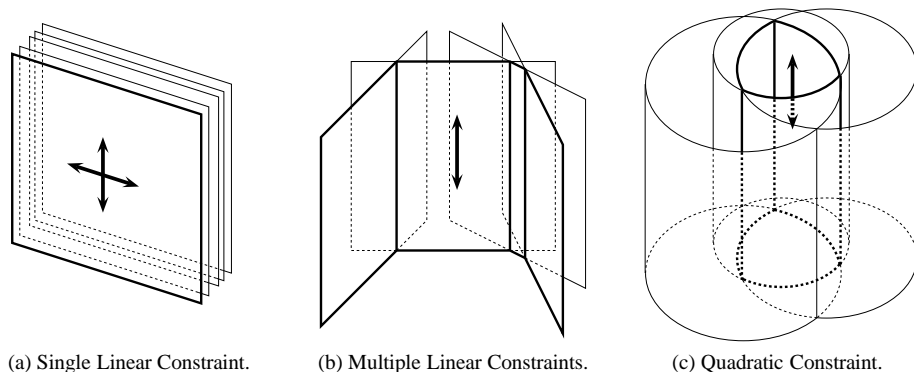


Figure 4.1: Illustration of Example 4.4 in \mathbb{R}^3 . The arrows indicate the dimension of the unconstrained subspace, equal to 3 minus the respective s-ranks α, β, γ .

After this intuitive example, the s-rank is now introduced in a rigorous manner for a fixed stage $i \in \mathbb{N}_1^N$ of (2.3). For every point $x \in \mathbb{X}$ and every value $\delta \in \Delta$, denote the corresponding level set of $f_i : \mathbb{R}^d \rightarrow \mathbb{R}$ by

$$F_i(x, \delta) := \{ \xi \in \mathbb{R}^d \mid f_i(x + \xi, \delta) = f_i(x, \delta) \} . \quad (4.8)$$

Let \mathcal{L} be the collection of all linear subspaces in \mathbb{R}^d . In order to be unconstrained, only those subspaces should be selected that are contained in all level sets $F_i(x, \delta)$:

$$\mathcal{L}_i := \bigcap_{\delta \in \Delta} \bigcap_{x \in \mathbb{X}} \{ L \in \mathcal{L} \mid L \subset F_i(x, \delta) \} . \quad (4.9)$$

More precisely, in (4.9) any Pr-null set could be removed from Δ that adversely influences the subsequent definition of the support rank; however this is not made explicit here to avoid an unnecessary complication of the notation.

Proposition 4.5 (Well-Definition of Unconstrained Subspace) *The collection \mathcal{L}_i contains a unique maximal element S_i in the set-inclusion sense, i.e. S_i contains all other elements of \mathcal{L}_i as subsets.*

Proof. First, note that \mathcal{L}_i is always non-empty, because for every $x \in \mathbb{X}$ and every $\delta \in \Delta$ the level set $F_i(x, \delta)$ includes the origin by its definition in (4.8), and therefore \mathcal{L}_i contains (at least) the trivial subspace $\{0\}$.

Second, introduce the partial order ' \preceq ' on \mathcal{L}_i defined by set inclusion; i.e. for any two elements $L_A, L_B \in \mathcal{L}_i$, $L_A \preceq L_B$ if and only if $L_A \subseteq L_B$. Since every chain in \mathcal{L}_i has an upper bound (namely \mathbb{R}^d), *Zorn's Lemma* (which is itself equivalent to the *Axiom of Choice*, see [7, p. 50]) implies that \mathcal{L}_i has at least one maximal element, in the set-inclusion ' \preceq ' sense.

Third, in order to prove that the maximal element is unique, suppose that L_A, L_B are both maximal elements of \mathcal{L}_i . It will be shown that $L_A \oplus L_B \in \mathcal{L}_i$, so that $L_A \neq L_B$ would contradict their maximality. According to (4.9), it must be shown that their direct sum $L_A \oplus L_B \subset F_i(x, \delta)$ for any fixed values $x \in \mathbb{X}$ and $\delta \in \Delta$. To see this, pick

$$\xi \in L_A \oplus L_B \implies \xi = \xi_A + \xi_B \quad \text{for } \xi_A \in L_A, \xi_B \in L_B ,$$

and then apply (4.8) twice to obtain

$$f_i(x + \xi_A + \xi_B, \delta) = f_i(x + \xi_A, \delta) = f_i(x, \delta) ,$$

because $\xi_A \in L_A$ and $\xi_B \in L_B$. □

Definition 4.6 (Unconstrained Subspace, Support Rank) (a) *The unique maximal element $S_i \in \mathcal{L}_i$ of Proposition 4.5 is called the unconstrained subspace of stage $i \in \mathbb{N}_1^N$.* (b) *The associated support rank (or s-rank) $\zeta'_i \in \mathbb{N}_0^d$ is defined as d minus the sub-space dimension of S_i ,*

$$\zeta'_i := d - \dim S_i .$$

Note that if \mathcal{L}_i contains only the trivial subspace, then the support rank actually equals to its upper bound d ; on the other hand, if \mathcal{L}_i contains more elements than the trivial subspace, then the support rank becomes strictly lower than d . The following theorem connects the s-rank to the s-dimension of a stage, proving that the s-rank of a stage represents an upper bound to its s-dimension.

Theorem 4.7 (Rank Condition) *Suppose that in (2.3) some stage $i \in \mathbb{N}_1^N$ has the s-rank $\zeta'_i \in \mathbb{N}_1^d$; then its s-dimension is bounded by $\zeta_i \leq \zeta'_i$.*

Proof. Without loss of generality, the proof is again given for the first stage $i = 1$. Pick any random multi-sample $\bar{\omega} \in \Delta^K$ (less any \Pr^K -null set for which the s-rank condition of Definition 4.6 may not hold).

By the assumption, there exists a (closed) linear subspace $S_1 \subset \mathbb{R}^d$ of dimension $d - \zeta'_1$ for which

$$f_1(x + \xi) = f_1(x) \quad \forall x, (x + \xi) \in \mathbb{X}, \xi \in S_1 .$$

Let S_1^\perp denote the orthogonal complement of S_1 ; recall that it is also a (closed) linear subspace in \mathbb{R}^d of dimension ζ'_1 and that every vector in \mathbb{R}^d can be uniquely written as the orthogonal sum of vectors in S_1 and S_1^\perp , e.g. [7, p. 135].

For the sake of a contradiction, suppose that $i = 1$ contributes more than ζ'_1 essential constraints to the resulting multi-stage RP (2.4). Denoting the samples that generate the essential constraints by

$$\tilde{\omega}_i := \{ \bar{\delta}^{(i, \kappa_i)} \mid \kappa_i \in \overline{\text{Es}}_i \} \quad \forall i = 1, \dots, N ,$$

this means that $\text{card}(\tilde{\omega}_1) \geq \zeta'_1 + 1$. It will be shown that this contradicts the optimality of the solution of (2.3),

$$\bar{x}_0^* := \bar{x}^*(\bar{\omega}^{(1)}, \dots, \bar{\omega}^{(N)}) = \bar{x}^*(\tilde{\omega}^{(1)}, \dots, \tilde{\omega}^{(N)}) .$$

Some additional notation is introduced to simplify the proof. Define

$$\tilde{\mathbb{X}} := \bigcap_{i \in \mathbb{N}_2^N} \bigcap_{\kappa_i \in \overline{\text{Es}}_i} \{ x \in \mathbb{X} \mid f_i(x, \bar{\delta}^{(\kappa_i)}) \leq 0 \}$$

as the feasible set with respect to all constraints, except for those pertaining to stage $i = 1$; being the intersection of convex sets, $\tilde{\mathbb{X}}$ is convex. Moreover for any $\kappa_1 \in \overline{\text{Es}}_1$, define

$$\bar{x}_{\kappa_1}^* := \bar{x}^*(\tilde{\omega}^{(1)} \setminus \{ \bar{\delta}^{(1, \kappa_1)} \}, \tilde{\omega}^{(2)}, \dots, \tilde{\omega}^{(N)})$$

as the new solutions when the essential constraint κ_1 is omitted from the reduced problem of (2.3). Recall that an essential constraint of (2.4) is a support constraint of the reduced problem by Definition 3.2(b), so the solution moves away from \bar{x}_0^* when a constraint $\kappa_1 \in \overline{\text{Es}}_1$ is omitted. Let the collection of all randomized solutions be written as

$$X := \{ \bar{x}_{\kappa_1}^* \mid \kappa_1 \in \overline{\text{Es}}_1 \} \cup \{ \bar{x}_0^* \} .$$

Observe that each $\bar{x}_{\kappa_1}^*$, for $\kappa_1 \in \overline{\text{Es}}_1$, is feasible with respect to all constraints, except for the κ_1 -th one, which is necessarily violated according to Definition 3.1. In consequence, all elements in X are pairwise distinct.

Since \mathbb{R}^d is the orthogonal direct sum of S_1 and S_1^\perp , for each point in X there is a unique orthogonal decomposition

$$\bar{x}_{\kappa_1}^* = v_{\kappa_1} + w_{\kappa_1} , \quad \text{where } v_{\kappa_1} \in S_1, w_{\kappa_1} \in S_1^\perp , \quad \forall \kappa_1 \in \overline{\text{Es}}_1 \cup \{0\} .$$

Consider the set

$$W := \{ w_{\kappa_1} \mid \kappa_1 \in \overline{\text{Es}}_1 \cup \{0\} \}$$

containing at least $\zeta'_1 + 2$ points in the ζ'_1 -dimensional subspace S_1^\perp . Applying Radon's Theorem [26, p. 151], W can be split into two disjoint subsets W_A and W_B such that there exists a point \tilde{w} in the intersection of their convex hulls:

$$\tilde{w} \in \text{conv}(W_A) \cap \text{conv}(W_B) . \quad (4.10)$$

When the indices in $\overline{\text{Es}}_1 \cup \{0\}$ are split correspondingly into I_A and I_B , it can be observed that every point $w_A \in \text{conv}(W_A)$ satisfies the constraints

$$f_1(w_A, \bar{\delta}^{(1, \kappa_1)}) \leq 0 \quad \forall \kappa_1 \in I_B .$$

This is true because it holds for every element in W_A and hence, by convexity, for every element in its convex hull. For the same reason, every point $w_B \in \text{conv}(W_B)$ satisfies the constraints

$$f_1(w_B, \bar{\delta}^{(1, \kappa_1)}) \leq 0 \quad \forall \kappa_1 \in I_A .$$

Hence from (4.10) it follows that

$$f_1(\tilde{w}, \bar{\delta}^{(1, \kappa_1)}) \leq 0 \quad \forall \kappa_1 \in \overline{\text{Es}}_1 . \quad (4.11)$$

According to (4.10), \tilde{w} can be expressed as a convex combination of elements in W_A or W_B . Splitting the points in X into X_A and X_B correspondingly and applying the same convex combination yields some

$$\tilde{x} \in \text{conv}(X_A) \cap \text{conv}(X_B) , \quad (4.12)$$

and in consequence also some $\tilde{v} \in S_1$.

To establish the contradiction two things remains to be verified: first that \tilde{x} is feasible with respect to all constraints, and second that it has a lower cost (or a better tie-break value) than \bar{x}_0^* . For the first, $\tilde{x} \in \tilde{X}$ because all points of X are in the convex set \tilde{X} and \tilde{x} is in their convex hull. Moreover,

$$f_1(\tilde{x}, \bar{\delta}^{(1, \kappa_1)}) = f_1(\tilde{w}, \bar{\delta}^{(1, \kappa_1)}) \leq 0 \quad \forall \kappa_1 \in \overline{\text{Es}}_1$$

because of (4.11). For the second, take the set from X_A and X_B which does not contain \bar{x}_0^* ; without loss of generality, say this is X_A . By construction all elements of X_A have a strictly lower target function value (or at least a better tie-break value) than \bar{x}_0^* . By linearity this also holds for points in $\text{conv}(X_A)$, and so $\tilde{x} \in \text{conv}(X_A)$ by (4.12). \square

5 Randomized Approximation of MSP Solution

In this section, the central result regarding the approximation of the solution to the MSP by the solution of the multi-stage RP is stated and proven. The main theorem in the first part provides an implicit link between the sample sizes K_1, \dots, K_N and the probability measures of the events that $V_1 \leq \varepsilon_1, \dots, V_N \leq \varepsilon_N$. Based on this result, an explicit relationship is derived in the second part.

5.1 The Sampling Theorem

This entire section is concerned with the proof of the following key result.

Theorem 5.1 (Sampling Theorem) *Consider problem (2.3) under all of the previous assumptions. For every stage $i = 1, \dots, N$ it holds that*

$$\Pr^K [V_i(\omega^{(1)}, \dots, \omega^{(N)}) > \varepsilon_i] \leq \Phi(\zeta'_i - 1; K_i, \varepsilon_i) , \quad (5.1)$$

where ζ'_i denotes the s -rank of stage i and $\Phi(\cdot; \cdot, \cdot)$ the cumulative probability distribution of a binomial random variable, as defined in Appendix A.

Without loss of generality, it again suffices to prove the result for stage $i = 1$. As opposed to existing results [10, Thm. 3.3] and [12, Thm. 2.4], the proof has to account for the simultaneous presence of the sampled constraints of the other stages $i = 2, \dots, N$, as well as the possibility of degeneracy and non-multi-regularity. The main idea of the proof is to obtain an upper bound on the distribution of the random violation probability $V_1(\omega^{(1)}, \dots, \omega^{(N)})$ in (5.1); it is divided into four main parts in order to provide a better overview of its fundamental structure.

Preliminaries

As in Lemma 3.6, the number of sampled constraints for stage $i = 1$ shall be varied in a thought experiment, while the other sample sizes K_2, \dots, K_N maintain their constant values. The variable sample size $\tilde{K}_1 \in \mathbb{N}_{\zeta_1}^\infty$ is used to distinguish this thought experiment from the actual sample size K_1 , and the variable multi-sample $\tilde{\omega}^{(1)}$ to distinguish from the actual multi-sample $\omega^{(1)}$. The total number of samples is then denoted by $\tilde{K} := \tilde{K}_1 + K_2 + \dots + K_N$ and the corresponding multi-sample by $\tilde{\omega}$.

The lack of multi-regularity is accounted for by considering only those cases of $z_1 := \text{card}(\text{Es}_1)$ that have a non-zero probability for any $\tilde{K}_1 \geq \zeta_1 + 1$. Let $Z_1 \subset \mathbb{N}_0^{\zeta_1}$ be the set of cardinalities of Es_1 which are associated with a non-zero probability measure,

$$Z_1 := \{z_1 \in \mathbb{N}_0^{\zeta_1} \mid \Pr^{\tilde{K}}[\text{card}(\text{Es}_1) = z_1] > 0\} . \quad (5.2)$$

According to Lemma 3.6, Z_1 is indeed well-defined as it is independent of any choice of the sample size \tilde{K}_1 which is admissible by Assumption 4.3. Hence it makes sense to consider a fixed value $\bar{z}_1 \in Z_1$ from here on, while keeping the sample size $\tilde{K}_1 \geq \bar{z}_1 + 1$ as a variable.

For any $\tilde{K}_1 \in \mathbb{N}_{\bar{z}_1+1}^\infty$, let $E \subset \mathbb{N}_1^{\tilde{K}_1}$ be an arbitrary subset of cardinality \bar{z}_1 that contains the indices of \bar{z}_1 specific choices from the \tilde{K}_1 sampled constraints. Define \mathcal{E}_E as the event where E includes exactly all of the \bar{z}_1 essential constraints contributed to (2.3) by stage $i = 1$, i.e. where $E = \text{Es}_1$. The conditional probability of the event \mathcal{E}_E , given that $z_1 = \bar{z}_1$,

$$\Pr_{z_1=\bar{z}_1}^{\tilde{K}}[\mathcal{E}_E] := \Pr_{z_1=\bar{z}_1}^{\tilde{K}}\{\tilde{\omega} \in \Delta^{\tilde{K}} \mid \mathcal{E}_E\} , \quad (5.3)$$

is now computed in two ways and the results are then equated. Note that both computations leave the hypothetical sample size \tilde{K}_1 open, and hence they work for any selection of the sample size within its allowed range of $\mathbb{N}_{\bar{z}_1+1}^\infty$.

Computation of the conditional probability

The first way of computing the conditional probability in (5.3) is an immediate consequence of the fact that each subset of $\mathbb{N}_1^{\tilde{K}_1}$ with cardinality \bar{z}_1 is equally likely to constitute the minimal essential set Es_1 . This is true because all \tilde{K}_1 sampled constraints in (2.3b) for stage $i = 1$ are generated by i.i.d. random variables (Assumption 2.4 and Section 3.2) and are thus identical from a probabilistic point of view. Given that $z_1 = \bar{z}_1$, there exist a total of \tilde{K}_1 choose \bar{z}_1 possible choices for Es_1 , so it follows that

$$\Pr_{z_1=\bar{z}_1}^{\tilde{K}}[\mathcal{E}_E] = \binom{\tilde{K}_1}{\bar{z}_1}^{-1} . \quad (5.4)$$

For the second way of computing the conditional probability in (5.3), observe that, given that $\text{card}(\text{Es}_1) = \bar{z}_1$, the *event* \mathcal{E}_E occurs if and only if all of the $\tilde{K}_1 - \bar{z}_1$ constraints in $E^C := \mathbb{N}_1^{\tilde{K}_1} \setminus E$ are not in the minimal essential set Es_1 . Given that $\text{card}(\text{Es}_1) = \bar{z}_1$, the *probability* of \mathcal{E}_E is therefore equal to the probability that all constraints in E^C are not essential constraints. To compute this probability, one may start by considering the specific case in which (2.3) includes only the constraints in E , i.e.

$$\text{MRP}[\{\delta^{(1,\kappa_1)} \mid \kappa_1 \in E\}, \omega^{(2)}, \dots, \omega^{(N)}] .$$

Given that $\text{card}(\text{Es}_1) = \bar{z}_1$, the conditional probability that $\text{Es}_1 = E$ is one. Now extract another random sample $\delta^{(1,\bar{z}_1+1)}$ for stage $i = 1$ and let v be the probability that this constraint becomes essential. Note that v is itself a random variable since it depends, in particular, on the outcome of the random samples $\delta^{(1,\kappa_1)}$ with $\kappa_1 \in E$, as well as the multi-samples of the other stages $\omega^{(2)}, \dots, \omega^{(N)}$ (and potentially their corresponding tie-break values due to Definition 3.4).

The event that the newly sampled constraint does not become essential has the probability $(1 - v)$, in which case the essential set remains $\text{Es}_1 = E$. In this case, the previous procedure can be repeated for

another sample $\delta^{(1, \bar{z}_1+2)}$, for which all of the above holds in the same way. Therefore, in terms of v , the probability that *none* of the additionally extracted $\tilde{K}_1 - \bar{z}_1$ sampled constraints become essential is equal to $(1 - v)^{(\tilde{K}_1 - \bar{z}_1)}$.

Even though the probability distribution of v is unknown, some distribution function $F_{\bar{z}_1} : [0, 1] \rightarrow [0, 1]$ can be introduced as a placeholder. It allows to express the conditional probability that, after extracting \tilde{K}_1 samples, E contains exactly all constraints in Es_1 , given that $z_1 = \bar{z}_1$, as

$$\Pr_{z_1=\bar{z}_1}^{\tilde{K}}[\mathcal{E}_E] = \int_0^1 (1 - v)^{(\tilde{K}_1 - \bar{z}_1)} dF_{\bar{z}_1}(v) . \quad (5.5)$$

Upper bound on the distribution

Since (5.4) and (5.5) express the same quantity, computed from different perspectives, they can be equated. This yields a set of integral equations

$$\int_0^1 (1 - v)^{(\tilde{K}_1 - \bar{z}_1)} dF_{\bar{z}_1}(v) = \left(\frac{\tilde{K}_1}{\bar{z}_1} \right)^{-1} \quad \forall \tilde{K}_1 \in \mathbb{N}_{\bar{z}_1}^\infty \quad (5.6)$$

which must be satisfied by the unknown probability distribution function $F_{\bar{z}_1}$. Here the missing condition for $\tilde{K}_1 = \bar{z}_1$ follows trivially from the property that any probability distribution function has integral one.

Solving (5.6) for $F_{\bar{z}_1}$ is known in the literature as a *Hausdorff Moment Problem* [25]. If a solution can be found, it is necessarily unique, as shown by [24, Cor. II.12.1] or [5, Thm. 30.1]. A comparison with the Euler Integral of the first kind [1, 6.2.1 and 6.2.2] immediately reveals that the following choice for probability density function solves (5.6), from which the desired (unique) distribution function can hence be deduced:

$$\frac{dF_{\bar{z}_1}(v)}{dv} = \bar{z}_1 v^{\bar{z}_1-1} \quad \implies \quad F_{\bar{z}_1}(v) = v^{\bar{z}_1} . \quad (5.7)$$

Completing the proof

Now consider again the actual sample size, which satisfies $K_1 \geq \zeta_1 + 1 \geq \bar{z}_1 + 1$ (by Assumption 4.3), and the event $\mathcal{E}_{E, \varepsilon_1}$ in which two things hold: (a) as in \mathcal{E}_E , a specific subset $E \subset \mathbb{N}_1^{K_1}$ of $\text{card}(E) = \bar{z}_1$ constitutes the minimal essential set Es_1 , and (b) the probability that a newly extracted constraint becomes an essential constraint of the augmented problem is greater than ε_1 .

With the distribution function $F_{\bar{z}_1}$ from (5.7), the conditional probability of $\mathcal{E}_{E, \varepsilon_1}$, given that $z_1 = \bar{z}_1$, becomes

$$\begin{aligned} \Pr_{z_1=\bar{z}_1}^{\tilde{K}}[\mathcal{E}_{E, \varepsilon_1}] &= \int_{\varepsilon_1}^1 (1 - v)^{(K_1 - \bar{z}_1)} dF_{\bar{z}_1}(v) \\ &= \bar{z}_1 \int_{\varepsilon_1}^1 v^{\bar{z}_1-1} (1 - v)^{(K_1 - \bar{z}_1)} dv \\ &= \bar{z}_1 B(1 - \varepsilon_1; K_1 - \bar{z}_1 + 1, \bar{z}_1) \\ &= \left(\frac{K_1}{\bar{z}_1} \right)^{-1} \Phi(\bar{z}_1 - 1; K_1, \varepsilon_1) , \end{aligned} \quad (5.8)$$

where identity (A.6) from Appendix A has been used to transform the incomplete beta function $B(\cdot; \cdot, \cdot)$ obtained by (A.4) into the binomial distribution function $\Phi(\cdot; \cdot, \cdot)$.

In order to bring the result in (5.8) closer to the claim, consider the new event $\mathcal{E}_{\bar{z}_1, \varepsilon_1}$ in which (a) Es_1 is now an arbitrary subset $E \subset \mathbb{N}_1^{K_1}$ of $\text{card}(E) = \bar{z}_1$, and (b) the probability that a newly extracted constraint becomes an essential constraint of the augmented problem is greater than ε_1 . By the same line of argument as above and using (5.8),

$$\begin{aligned} \Pr_{z_1=\bar{z}_1}^K[\mathcal{E}_{\bar{z}_1, \varepsilon_1}] &= \left(\frac{K_1}{\bar{z}_1} \right) \Pr_{z_1=\bar{z}_1}^K[\mathcal{E}_{E, \varepsilon_1}] \\ &= \Phi(\bar{z}_1 - 1; K_1, \varepsilon_1) . \end{aligned} \quad (5.9)$$

Finally, the Sampling Lemma 3.5 has shown that the fact that a newly extracted constraint violates the current solution implies that it must be an essential constraint of the augmented problem. Therefore the probability of violating the current solution is always smaller than or equal to the probability of becoming an essential constraint of the augmented problem, i.e.

$$\Pr_{z_1=\bar{z}_1}^K [V_1(\omega^{(1)}, \dots, \omega^{(N)}) > \varepsilon_1] \leq \Pr_{z_1=\bar{z}_1}^K [\mathcal{E}_{\bar{z}_1, \varepsilon_1}] .$$

Moreover, for each $\bar{z}_1 \in Z_1$, as defined in (5.2), it holds that

$$\Pr_{z_1=\bar{z}_1}^K [\mathcal{E}_{\bar{z}_1, \varepsilon_1}] = \Phi(\bar{z}_1 - 1; K_1, \varepsilon_1) \leq \Phi(\zeta'_1 - 1; K_1, \varepsilon_1) ,$$

from the monotonicity property of the binomial distribution function that follows from its definition in (A.1). Given that the events of $\bar{z}_1 \in Z_1$ are collectively exhaustive and mutually exclusive,

$$\begin{aligned} \Pr^K [V_1(\omega) > \varepsilon_1] &= \sum_{\bar{z}_1 \in Z_1} \Pr_{z_1=\bar{z}_1}^K \{ \omega \in \Delta^K \mid V_1(\omega) > \varepsilon_1 \} \cdot \Pr^K [z_1 = \bar{z}_1] \\ &\leq \sum_{\bar{z}_1 \in Z_1} \Pr_{z_1=\bar{z}_1}^K \{ \omega \in \Delta^K \mid \mathcal{E}_{\bar{z}_1, \varepsilon_1} \} \cdot \Pr^K [z_1 = \bar{z}_1] \\ &\leq \sum_{\bar{z}_1 \in Z_1} \Phi(\zeta'_1 - 1; K_1, \varepsilon_1) \cdot \Pr^K [z_1 = \bar{z}_1] \\ &= \Phi(\zeta'_1 - 1; K_1, \varepsilon_1) . \end{aligned} \tag{5.10}$$

This completes the proof of Theorem 5.1.

5.2 Explicit Bounds on the Sample Sizes

Formula (5.1) in Theorem 5.1 ensures a *confidence level* of $1 - \Phi(\zeta'_i - 1; K_i, \varepsilon_i)$ that the violation probability $V_i(\omega^{(1)}, \dots, \omega^{(N)}) \leq \varepsilon_i$, based on the sample K_i for stage i . However, in practical applications a given confidence level $(1 - \theta_i) \in (0, 1)$ is usually imposed, for which an appropriate sample size K_i needs to be identified.

The most accurate way of finding this sample size is by observing that $\Phi(\zeta'_i - 1; K_i, \varepsilon_i)$ is a monotonically decreasing function in K_i and applying a numerical procedure (e.g. based on bisection) for computing the smallest sample size that ensures $\Phi(\zeta'_i - 1; K_i, \varepsilon_i) \leq \theta_i$. The resulting K_i shall be referred to as the *implicit bound* on the sample size.

For a qualitative analysis of the behavior of this implicit bound for changing values of ε_i and θ_i (and also for initialization of a bisection-based algorithm), it is useful to derive an *explicit bound* on the sample size K_i . Since formula (5.1) cannot be readily inverted, the binomial distribution function must first be controlled by some upper bound, which is then inverted. The details of this procedure are well-documented in the literature and therefore omitted.

A common approach using a Chernoff bound [14], as shown in [9, Rem. 2.3] and [10, Sec. 5], provides a simple explicit formula for K_i :

$$K_i \geq \frac{2}{\varepsilon_i} \left[\log\left(\frac{1}{\theta_i}\right) + \zeta'_i - 1 \right] , \tag{5.11}$$

where $\log(\cdot)$ denotes the natural logarithm. As shown in [2, Cor. 1], this bound can be further improved to a more complicated formula for K_i :

$$K_i \geq \frac{1}{\varepsilon_i} \left[\log\left(\frac{1}{\theta_i}\right) + \sqrt{2(\zeta'_i - 1) \log\left(\frac{1}{\theta_i}\right)} + \zeta'_i - 1 \right] . \tag{5.12}$$

6 The Sampling-and-Discarding Approach

The sampling-and-discarding approach has been examined in the context of single-stage RPs [10, 13]; it can also be extended to multi-stage RPs as shown in this section. The underlying goal is to improve the quality of the randomized solution, i.e. its objective value, while keeping the probability of violating the nominal chance constraints under control. To this end, the sample sizes K_i are deliberately increased above the bounds derived in Section 5, in exchange for allowing a certain number of R_i sampled constraints to be discarded *ex-post*, that is after the outcomes of the samples have been observed.

Appropriate discarding procedures are introduced in the first part of this section. Then the main result of this section is stated, providing an implicit formula for selecting the right pairs of sample size K_i and discarded constraints R_i ; these may differ again between stages, in particular a non-removal strategy may still be followed for some of stages i (by putting $R_i = 0$). Finally, explicit bounds for the choice of pairs K_i and R_i are indicated, in analogy to Section 5.

6.1 Constraint Discarding Procedure

For each stage of the problem, the discarding procedure is performed by an arbitrary, pre-defined (*sample*) *removal algorithm*.

Definition 6.1 (Removal Algorithm) *For each stage $i = 1, \dots, N$ of (2.3), the (sample) removal algorithm $\mathcal{A}_i^{(K_i, R_i)} : \Omega \rightarrow \tilde{\Omega}_i$ is a deterministic function of the multi-samples of all stages, $\omega \in \Omega$. It returns a subset of samples $\tilde{\omega}_i \in \tilde{\Omega}_i$, in which R_i out of K_i samples have been removed from ω_i , i.e. $\text{card}(\tilde{\omega}_i) = K_i - R_i$.*

With view on improving the quality of the solution, of course, the algorithm should aim at lowering the objective function value $\text{MRP}[\tilde{\omega}^{(1)}, \dots, \tilde{\omega}^{(N)}]$ as much as possible. Various possibilities for removal algorithms are described in detail by [10, Sec. 5.1], and further references are provided by [13, Sec. 2]. Some brief description of important removal algorithms are given below.

Example 6.2 (a) Optimal Constraint Removal. The best improvement of the objective function value is achieved by solving the reduced problem for all possible combinations of discarding R_i out of K_i constraints. However, this generally leads to a combinatorial complexity of the optimal constraint removal algorithm, which becomes computationally intractable for larger values of R_i —in particular when constraints of multiple stages have to be discarded.

(b) Greedy Constraint Removal. Starting by solving the problem for all K_i constraints, a greedy constraint removal algorithm performs R_i consecutive steps, where in each step one single constraint is removed according the optimal procedure of (a). Different stages can also be handled consecutively, or optimally. For most practical problems this algorithm can be expected to work almost as good as (a), while posing a much lower computational burden.

(c) Marginal Constraint Removal. Similar to the algorithm in (b) the constraints are removed consecutively, however based on the largest marginal cost improvement (given by the corresponding Lagrange multiplier [8, Cha. 5]) in each step, instead of the largest total cost improvement. The marginal constraint removal algorithm can be designed to remove one constraint of each stage in a single step of the procedure, or to proceed consecutively among the stages.

Existing theory for single-stage RPs [10, Sec. 4.1.1] and [13, Ass. 2.2] assumes that all of the removed constraints are violated by the reduced randomized solution. While this assumption is also sufficient for the purposes of multi-stage RPs, it may turn out to be too restrictive in some problem instances. In fact, due to the interplay of multiple stages, it may be impossible to find R_i constraints that are violated by the randomized solution—a situation that may also occur in the case of a single stage and the presence of a deterministic constraint set \mathbb{X} . For this case the alternative Assumption 6.3(b) of the *monotonicity property* is offered, which is explained below.

Assumption 6.3 (Stage with Discarded Constraints) Every stage $i \in \mathbb{N}_1^N$ of (2.3) with $R_i > 0$ satisfies at least one of the following two conditions: (a) For almost every $\omega \in \Omega$, each of the constraints discarded by the removal algorithm $\mathcal{A}_i^{(K_i, R_i)}(\omega)$ is violated by the reduced solution:

$$f_i(x^*(\tilde{\omega}_1, \dots, \tilde{\omega}_N), \delta^{(i, \kappa_i)}) > 0 \quad \forall \delta^{(i, \kappa_i)} \in \omega \setminus \tilde{\omega} .$$

(b) The stage satisfies the monotonicity property of Definition 6.4 below.

Definition 6.4 (Monotonicity Property) A stage $i \in \mathbb{N}_1^N$ of (2.3) enjoys the monotonicity property if for every $K_i \in \mathbb{N}$ and almost every $\omega^{(i)} \in \Delta^{K_i}$ the following condition holds: each point in the sampled feasible set of stage i ,

$$\mathbb{X}_i(\omega^{(i)}) := \{\xi \in \overline{\mathbb{R}}^d \mid f_i(\xi, \delta^{(i, \kappa_i)}) \leq 0 \quad \forall \kappa_i \in \mathbb{N}_1^{K_i}\} \quad (6.1)$$

where $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$, is violated by a new sampled constraint only if also the cost-minimal point in $\mathbb{X}_i(\omega^{(i)})$,

$$x_i^*(\omega^{(i)}) := \arg \min \{c^T \xi \mid \xi \in \mathbb{X}_i(\omega^{(i)})\} \quad (6.2)$$

is violated. More precisely, for each $\xi \in \mathbb{X}_i(\omega^{(i)})$ and almost every $\delta \in \Delta$,

$$f_i(\xi, \delta) > 0 \quad \implies \quad f_i(x_i^*(\omega^{(i)}), \delta) > 0 . \quad (6.3)$$

At first glance, Definition (6.4) appears to be abstract, yet it is easy to check for most practical problems without involving any calculations. The following example provides the necessary intuition.

Example 6.5 (Monotonic and Non-Monotonic Stages) Consider the MRP of (2.3) in $d = 2$ dimensions, where $\mathbb{X} = [-10, 10]^2 \subset \mathbb{R}^2$, the target function vector is $c = [0 \ 1]^T$, and the number of stages is $N = 2$. For testing the monotonicity of each stage $i = 1, 2$, problem (2.3) is considered as a single-stage RP, i.e. the compact set \mathbb{X} and the respective other stage are neglected.

(a) *Monotonic Stage.* Suppose the first stage $i = 1$ is of the linear form

$$\begin{bmatrix} \delta_1^{(1, \kappa_1)} & 1 \end{bmatrix} x - \delta_2^{(1, \kappa_1)} \leq 0 \quad \forall \kappa_1 = 1, \dots, K_1 ,$$

where $\delta_1^{(1, \kappa_1)} \in \{-1, 1\}$ and $\delta_2^{(1, \kappa_1)} \in [-1, 1]$. Observe that for any number of samples $K_1 \in \mathbb{N}$ and any sample outcomes, the additional sample δ_1, δ_2 either cuts off no point from the feasible set of stage $i = 1$, $\mathbb{X}_1(\omega^{(1)})$, or the cut-off set includes the cost-minimal point $x_1^*(\{\delta^{(1, 1)}, \dots, \delta^{(1, K_1)}\})$. Therefore stage $i = 1$ is *monotonic* in the sense of Definition 6.4. This argument is illustrated in Figure 6.1(a).

(b) *Non-Monotonic Stage.* Suppose the second stage $i = 2$ is of the linear form

$$\begin{bmatrix} \delta_3^{(2, \kappa_2)} & 1 \end{bmatrix} x - \delta_4^{(2, \kappa_2)} \leq 0 \quad \forall \kappa_2 = 1, \dots, K_2 ,$$

where $\delta_3^{(2, \kappa_2)} \in [-1, 1]$ and $\delta_4^{(2, \kappa_2)} \in [-1, 1]$. Observe that for any $K_2 \in \mathbb{N}$ it is possible for a new sample δ_3, δ_4 to cut off some previously feasible point from $\mathbb{X}_2(\omega^{(2)})$, without removing also the cost-minimal point $x_2^*(\{\delta^{(2, 1)}, \dots, \delta^{(2, K_2)}\})$. A possible configuration of this type is shown in Figure 6.1(b). Therefore stage $i = 2$ is *not monotonic* in the sense of Definition 6.4.

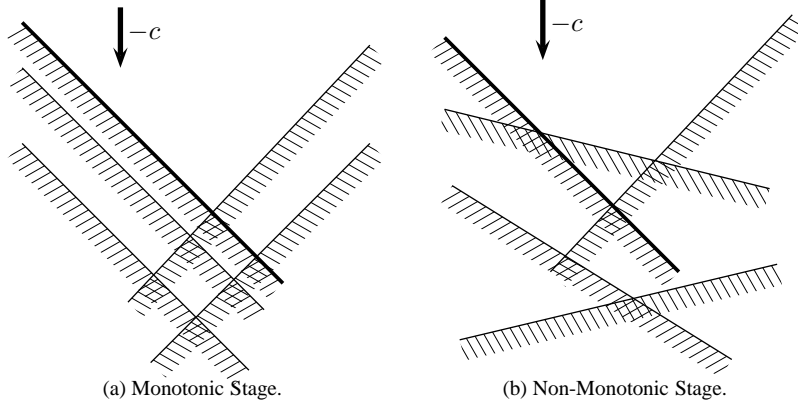


Figure 6.1: Illustration of Example 6.5. Non-bold constraints are generated by the multi-sample $\omega^{(i)} \in \Delta^{K_i}$ of stage $i = 1, 2$; bold constraints result from the actual uncertainty $\delta \in \Delta$. Figure (a) depicts a stage for which it is not possible for a new sample to cut off a feasible point without removing also the optimum; Figure (b) depicts a case where a feasible point is indeed cut off without removing the optimum.

The usefulness of the monotonicity property for RPs is revealed by the following result, whose proof is an immediate consequence of Definition 6.4 and therefore omitted.

Lemma 6.6 *If a stage $i \in \mathbb{N}_1^N$ of (2.3) enjoys the monotonicity property, then for every $K_i \in \mathbb{N}$ and almost every $\omega^{(i)} \in \Delta^{K_i}$:*

$$\Pr[f_i(\xi, \delta) > 0] \leq \Pr[f_i(x_i^*(\omega^{(i)}), \delta) > 0] \quad \forall \xi \in \mathbb{X}_i(\omega^{(i)}) . \quad (6.4)$$

In other words, with probability one every point ξ in its sampled feasible set $\mathbb{X}_i(\omega^{(i)})$ has a violation probability less than or equal to that of the cost-minimal point $x_i^(\omega^{(i)})$.*

6.2 The Discarding Theorem

For the sampling-and-discarding approach, the following result holds for the case of multi-stage RPs.

Theorem 6.7 (Discarding Theorem) *Consider problem (2.3) together with some ex-post discarding algorithms $\mathcal{A}_i^{(K_i, R_i)}$ returning the reduced multi-samples $\tilde{\omega}^{(i)}$ for the stages $i = 1, \dots, N$. Then it holds that*

$$\Pr^K[V_i(\tilde{\omega}^{(1)}, \dots, \tilde{\omega}^{(N)}) > \varepsilon_i] \leq \binom{R_i + \zeta_i' - 1}{R_i} \Phi(R_i + \zeta_i' - 1; K_i, \varepsilon_i) , \quad (6.5)$$

where ζ_i' denotes the support rank of stage i and $\Phi(\cdot; \cdot, \cdot)$ the binomial distribution as defined in Appendix A.

Proof. Here the multi-stage RP case is reduced to the single-stage RP case, for which two detailed proofs are available in [10, Sec. 4.1.1] and [13, Sec. 5.1].

In particular, suppose that Assumption 6.3(a) holds. The proof in [13, Sec. 5.1] works analogously for an arbitrary stage $i \in \mathbb{N}_1^N$, given that an upper bound of the violation distribution is readily available from the proof of Theorem 5.1.

Alternatively, suppose that Assumption 6.3(b) holds. In this case the proof in [13, Sec. 5.1] can be applied directly to the single-stage problem that arises from (2.3) when all other stages and \mathbb{X} are omitted. In consequence, (6.5) holds for the cost-minimal point of this single-stage RP. But given that the stage is monotonic, by Lemma 6.6 it also holds for any feasible point, in particular for the optimal point of (2.3). \square

An excellent account of the merits of the sampling-and-discarding approach is provided in [13], and therefore it need not be restated here. It should be emphasized, however, that the randomized solution converges to the true solution of the MSP as the number of discarded constraints increases, given the constraints are removed by the optimal algorithm in Example 6.2(a).

6.3 Explicit Bounds on the Sample-and-Discarding Pairs

Similar to multi-stage RPs without ex-post constraint removal in Section 5, explicit bounds on the sample size K_i can also be derived for the sampling-and-discarding approach, assuming the number of discarded constraints R_i to be fixed. The technicalities, using Chernoff bounds [14], have already been worked out for single-stage RPs; refer to [10, Sec. 5] for details. The resulting explicit bound becomes

$$K_i \geq \frac{2}{\varepsilon_i} \log\left(\frac{1}{\theta_i}\right) + \frac{4}{\varepsilon_i} (R_i + \zeta'_i - 1) , \quad (6.6)$$

where $\log(\cdot)$ denotes the natural logarithm. Similar techniques can be applied for obtaining an explicit bound on the number of discarded constraints R_i , assuming the sample size K_i to be fixed. The mathematical details and a further discussion are found in [13, Sec. 4.3], and the resulting bound is given by

$$R_i \leq \varepsilon_i K_i - \zeta'_i + 1 - \sqrt{2\varepsilon_i K_i \log\left(\frac{(\varepsilon_i K_i)^{\zeta'_i - 1}}{\theta_i}\right)} . \quad (6.7)$$

7 Example: Minimal Diameter Cuboid

In this final section the application of multi-stage RPs is demonstrated for an academic example, which has been selected to emphasize the potential of randomization in stochastic optimization in general and the advantage of multi-stage over single-stage RPs in particular problems.

7.1 Problem Statement

Let δ be an uncertain point in $\Delta \subset \mathbb{R}^n$, where both its distribution and support set may be unknown. The objective of this example is to construct the Cartesian product of closed intervals in \mathbb{R}^n (“ n -cuboid”) C of minimal n -diameter W , while being large enough to contain the uncertain point δ in its i -th coordinate with probability $1 - \varepsilon_i$; see Figure 7.1 for an illustration. The task resembles the problem of finding the smallest ball containing an uncertain point [15, Sec. 2], except that here the coordinates are constrained independently.

Denote by $z \in \mathbb{R}^n$ the center point of the cuboid and by $w \in \mathbb{R}_+^n$ the interval widths in each dimension, so that

$$C = \{\xi \in \mathbb{R}^n \mid |\xi_i - z_i| \leq w_i/2\} . \quad (7.1)$$

Then the corresponding stochastic program reads as follows,

$$\min_{z \in \mathbb{R}^n, w \in \mathbb{R}_+^n} \|w\|_2 , \quad (7.2a)$$

$$\text{s.t.} \quad \Pr[z_i - w_i/2 \leq \delta_i \leq z_i + w_i/2] \geq (1 - \varepsilon_i) \quad \forall i \in \mathbb{N}_1^n . \quad (7.2b)$$

Since the objective function is not linear, problem (7.2) first has to be rewritten by an epigraph reformulation (compare Remark 2.2(a)) into

$$\min_{z \in \mathbb{R}^n, w \in \mathbb{R}_+^n, W \in \mathbb{R}} W , \quad (7.3a)$$

$$\text{s.t.} \quad \|w\|_2 \leq W , \quad (7.3b)$$

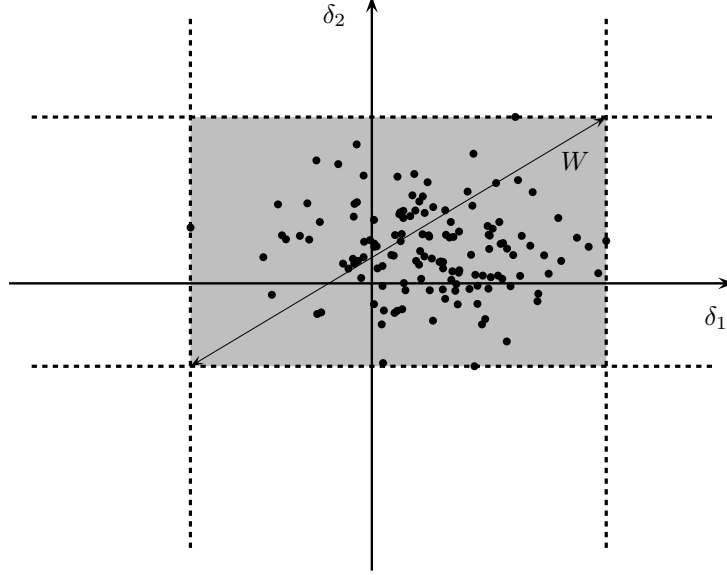


Figure 7.1: Illustration of the numerical example for \mathbb{R}^2 . The point $\delta \in \Delta$ appears at random in \mathbb{R}^2 , according to some unknown distribution; the points drawn here represent 166 i.i.d. samples of δ . The objective is to construct the smallest product of two closed intervals ('2-cuboid'), indicated here by the shaded rectangle, such that the probability of missing δ is smaller than $\varepsilon_1, \varepsilon_2$ in the respective dimension 1, 2.

$$\Pr \left[\max \{ z_i - w_i/2 - \delta_i, -z_i - w_i/2 + \delta_i \} \leq 0 \right] \geq (1 - \varepsilon_i) \quad (7.3c)$$

$$\forall i \in \mathbb{N}_1^n.$$

Note that (7.3) has the form of an MSP (2.1), for a $d = 2n + 1$ dimensional search space and $N = n$ stages. In particular, the objective function (7.3a) is linear, as it is a scalar; constraint (7.3b) is deterministic and convex; and each of the n stages (7.3c) contains a nominal chance constraints that is convex in z, v for any fixed value of the uncertainty $\delta \in \Delta$; see [8, Cha. 3]. Each of the stages $i = 1, \dots, n$ depends on exactly one component δ_i of the uncertainty, which is a special case of dependence on the entire uncertainty vector δ (compare Remark 2.2(c)). The convex and compact set \mathbb{X} is constructed from the positivity constraints on w , the deterministic and convex constraint (7.3b), and some very high artificial bounds introduced for all variables. Existence of a feasible solution, and hence the satisfaction of Assumption 2.3, is intuitively clear from the problem setup.

7.2 Randomization Approach

To solve (7.3) by the randomization approach, observe that each of the stages $i = 1, \dots, n$ has support rank $\zeta'_i = 2$ in the MRP corresponding to (7.3), as it only depends on the two variables z_i and w_i . For a fixed confidence level, e.g. $\theta = 10^{-6}$, the implicit sample sizes K_1, \dots, K_n in (5.1) can be computed for a given problem dimension n and chance constraint levels $\varepsilon_1, \dots, \varepsilon_n > 0$ by a bisection-based algorithm (see Section 5.2). For simplicity, the chance constraint levels are selected as equal for all stages, so that the implicit sample sizes are all identical.

Given the outcomes of all random multi-samples $\bar{\omega}^{(1)}, \dots, \bar{\omega}^{(n)}$, the randomized program is easily solved by the smallest n -cuboid that contains all sampled points; see also Figure 7.1. In other words, the resulting $\overline{\text{MRP}}$ can be determined analytically and without the use of any iterative optimization scheme. Table 7.1(a) summarizes the minimal sample sizes required for guaranteeing various chance constraint levels ε_i

in various dimensions n (with confidence $\theta = 10^{-6}$). The indicated numbers represent the (identical) implicit sample size for all stages $i = 1, \dots, n$, rather than the ones obtained by explicit bounds (5.11) or (5.12).

One can compare these sample sizes to those resulting from the single-stage RP theory (Table 7.1), and also the corresponding solution quality (Table 7.2). In order to apply the single-stage RP theory, first all of the (nominal) constraints in (7.3b,c) are combined into a single stage by (a) taking its constraint function as the point-wise maximum of the constraint functions $i = 1, \dots, n$ in (7.3c), and (b) taking the constraint level as $\min\{\varepsilon_1, \dots, \varepsilon_n\}$. It is easy to realize the conservatism introduced by this procedure (as described in more detail in Section 1).

sample size K_i	cuboid dimension $n =$						
	2	3	5	10	50	100	500
$\varepsilon_i =$ 1%	1,734	1,777	1,831	1,903	2,072	2,144	2,311
5%	341	349	360	374	407	421	454
10%	166	170	176	182	199	205	221
25%	62	63	65	67	73	76	82

(a) Multi-Stage Randomized Program.

sample size K_i	cuboid dimension $n =$						
	2	3	5	10	50	100	500
$\varepsilon_i =$ 1%	2,334	2,722	3,431	5,020	15,588	27,535	115,786
5%	459	536	677	992	3,095	5,477	23,093
10%	225	263	332	488	1,533	2,719	11,506
25%	84	99	125	186	595	1,063	4,550

(b) Single-Stage Randomized Program.

Table 7.1: Implicit sample sizes $K_1 = \dots = K_n$ for the multi-stage and the single-stage RP, based on a confidence level of $\theta = 10^{-6}$, depending on problem dimension n and chance constraint levels $\varepsilon_1 = \dots = \varepsilon_n$.

As a result, the sample sizes based on the single-stage approach are significantly larger than those of the multi-stage approach, as it can be seen from Table 7.1(b). In particular, it is not surprising that the multi-stage approach scales a lot better with growing dimensions n of the problem—the reason being that the support dimension of each stage does not grow with n , in contrast to Helly’s dimension. The remaining small growth of the sample size with the dimension in Table 7.1(a) owes to the fact that in the multi-stage approach the confidence level θ needs to be (evenly) distributed among the stages, i.e. $\theta_i = \theta/n$ for all $i = 1, \dots, n$, in order to allow for a fair comparison to the single-stage approach using θ .

A larger sample size increases the data requirements and computation effort, but also diminishes the quality of the randomized solution. The latter effect is quantified in Table 7.2, where the objective function value achieved by the single-stage RP is compared to that of the multi-stage RP (for the sample sizes in Table 7.1). The indicated values represent the averages over one million RPs, using a multi-variate standard normal distribution for the uncertainty δ .

relative obj. value	cuboid dimension $n =$						
	2	3	5	10	50	100	500
$\varepsilon_i =$ 1%	2.4%	3.4%	5.0%	7.5%	14.8%	18.4%	26.9%
5%	3.3%	4.6%	6.6%	9.8%	18.9%	23.8%	34.4%
10%	3.9%	5.4%	7.6%	11.5%	22.2%	27.4%	39.3%
25%	5.0%	7.2%	10.1%	15.1%	28.5%	34.7%	49.1%

Table 7.2: Objective function value of the single-stage case as relative surplus of the multi-stage case, based on the sample sizes in Table 7.1 and a multi-variate standard normal distribution for δ . Each of the indicated values represents an average over one million randomized solutions.

Acknowledgement

The research of G. Schildbach has received funding from the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement number FP7-ICT-2009-4 248940.

The research of L. Fagiano has received funding from the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement number PIOF-GA-2009-252284, Marie Curie project ‘Innovative Control, Identification and Estimation Methodologies for Sustainable Energy Technologies’.

A Probability Distributions

Several basic probability-related functions are used throughout this paper. The *Binomial Distribution Function* [1, p. 26.1.20]

$$\Phi(x; K, \varepsilon) := \sum_{j=0}^x \binom{K}{j} \varepsilon^j (1 - \varepsilon)^{K-j} \quad (\text{A.1})$$

expresses the probability of seeing at most $x \in \mathbb{N}_0^K$ successes in $K \in \mathbb{N}$ independent Bernoulli trials, where the probability of success is $\varepsilon \in (0, 1)$ per trial. The (real) *Beta Function* [1, p. 6.2.1]

$$B(a, b) := \int_0^1 \xi^{a-1} (1 - \xi)^{b-1} d\xi \quad (\text{A.2})$$

is defined for any parameters $a, b \in \mathbb{R}_+$, and $\xi \in (0, 1)$; it also satisfies the identity [1, p. 6.2.2]

$$B(a, b) = B(b, a) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \quad (\text{A.3})$$

where $\Gamma : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ denotes the (real) *Gamma Function* with $\Gamma(n+1) = n!$ for any $n \in \mathbb{N}_0^\infty$ [1, p. 6.1.5]. The corresponding *Incomplete Beta Function* [1, p. 6.6.1] is then given by

$$B(\varepsilon; a, b) := \int_0^\varepsilon \xi^{a-1} (1 - \xi)^{b-1} d\xi = \int_{1-\varepsilon}^1 \xi^{b-1} (1 - \xi)^{a-1} d\xi, \quad (\text{A.4})$$

where the last equality follows by a simple substitution. An important identity is obtained from [1, pp. 3.1.1, 6.6.2, 26.5.7],

$$B(\varepsilon; a, b) = B(a, b) \sum_{j=a}^{a+b-1} \binom{a+b-1}{j} \varepsilon^j (1 - \varepsilon)^{a+b-1-j}, \quad (\text{A.5})$$

which can be written more compactly by use of the binomial distribution (A.1), see for instance [10, p. 3437]:

$$B(\varepsilon; a, b) = \frac{1}{b} \binom{a+b-1}{b}^{-1} \Phi(b-1; a+b-1, 1 - \varepsilon). \quad (\text{A.6})$$

References

- [1] M. ABRAMOWITZ AND I.A. STEGUN, *Handbook of Mathematical Functions*, Dover Publications, New York, 9th ed., 1970.
- [2] T. ALAMO, R. TEMPO, AND A. LUQUE, *On the sample complexity of probabilistic analysis and design methods*, in *Perspectives in Mathematical System Theory, Control, and Signal Processing*, J.C Willems et al., ed., Springer, Berlin et al., 2010, pp. 39–50.
- [3] D. BAI, T. CARPENTER, AND J. MULVEY, *Making a case for robust optimization models*, *Management Science*, 43(7) (1997), pp. 895–907.
- [4] A. BEN-TAL AND A. NEMIROVSKI, *Robust convex optimization*, *Mathematics of Operations Research*, 23(4) (1998), pp. 769–805.
- [5] P. BILLINGSLEY, *Probability and Measure*, John Wiley & Sons, New York et al., 3rd ed., 1995.
- [6] J.R. BIRGE AND F. LOUVEAUX, *Introduction to Stochastic Programming*, Springer, New York, 1997.
- [7] B. BOLLOBÁS, *Linear Analysis*, Cambridge University Press, Cambridge et al., 2nd ed., 1999.
- [8] S. BOYD AND L. VANDENBERGHE, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.
- [9] G.C. CALAFIORE, *On the expected probability of constraint violation in sampled convex programs*, *Journal of Optimization Theory and Applications*, 143 (2009), pp. 405–412.
- [10] ———, *Random convex programs*, *SIAM Journal of Optimization*, 20(6) (2010), pp. 3427–3464.
- [11] G. CALAFIORE AND M.C. CAMPI, *Uncertain convex programs: Randomized solutions and confidence levels*, *Mathematical Programming, Series A*, 102-1 (2005), pp. 25–46.
- [12] M.C. CAMPI AND S. GARATTI, *The exact feasibility of randomized solutions of uncertain convex programs*, *SIAM Journal of Optimization*, 19 (2008), pp. 1211–1230.
- [13] ———, *A sampling and discarding approach to chance-constrained optimization: Feasibility and optimality*, *Journal of Optimization Theory and Applications*, 148 (2011), pp. 257–280.
- [14] H. CHERNOFF, *A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations*, *The Annals of Mathematical Statistics*, 23(4) (1952), pp. 493–507.
- [15] B. GÄRTNER AND E. WELZL, *A simple sampling lemma: Analysis and applications in geometric optimization*, *Discrete & Computational Geometry*, 25 (2001), pp. 569–590.
- [16] P. KALL AND J. MAYER, *Stochastic Linear Programming*, Springer, New York et al., 2nd ed., 2011.
- [17] P. KOUVELIS AND G. YU, *Robust Discrete Optimization and Its Applications*, Kluwer Academic Publishers, Dordrecht, 1997.
- [18] D. LUENBERGER AND Y. YE, *Linear and Nonlinear Programming*, Springer, Berlin et al., 3rd ed., 2008.
- [19] J.M. MULVEY AND R.J. VANDERBEI, *Robust optimization of large-scale systems*, *Operations Research*, 43(2) (1995), pp. 264–281.
- [20] J. NOCEDAL AND S.J. WRIGHT, *Numerical Optimization*, Springer, New York, 2nd ed., 2006.
- [21] A. PRÉKOPA, *Stochastic Programming*, Kluwer, Dordrecht et al., 1995.
- [22] R.T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, Princeton, 1970.

- [23] A. SHAPIRO, D. DENTCHEVA, AND A. RUSZCZYŃSKI, *Stochastic Programming, Modeling and Theory*, SIAM, Philadelphia, 2009.
- [24] A.N. SHIRYAEV, *Probability*, Springer, New York et al., 2nd ed., 1996.
- [25] J.A. SHOHAT AND J.D. TAMARKIN, *The Problem of Moments*, American Mathematical Society, Providence (RI), 4th ed., 1970.
- [26] G.M. ZIEGLER, *Lectures on Polytopes*, Springer, New York et al., 1st ed., 2007.